2016 Research, Science and Innovation Domain Plan

Ministry of Business, Innovation and Employment
Statistics New Zealand
Ministry of Education
Tertiary Education Commission

September 2016

MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

New Zealand Government

Foreword

International consensus is that innovation is critical to long-run economic growth, health outcomes, social development and environmental sustainability. Research and science expand our understanding and knowledge of the world we live in, which ultimately leads to real improvements in the quality of our lives. But innovation is not confined to the research lab. It goes beyond to users, suppliers and consumers everywhere in government, business and non-profit organisations.

Government intervention in the innovation system is well established by market failure principles and by empirical studies showing high rates of social return from research and development (R&D). Governments thus have a significant and enduring role to play in their innovation systems.

Over recent years government investment in R&D has increased significantly, with a commitment to growing investment to 0.8 per cent of GDP. This spending spans the primary industries, science and innovation, and tertiary education portfolios, but it also spills over to others such as health and environment.

The growing investment in and importance of the innovation system increase the need to better understand New Zealand's system, demonstrate results and impacts, increase transparency, raise efficiencies and reduce transaction costs. Policy settings and investment decisions need to be underpinned by sound evidence to ensure value for money. Data plays a critical part in generating that evidence and informing evaluation. We also recognise that making sense of data requires analytics and interpretation.

Data on research, science and innovation in New Zealand has suffered from a lack of coordination and oversight across government agencies. This has placed considerable burden on the research community and has impeded our ability to generate timely, quality and reliable data across the system. Furthermore, much of government agencies' relevant administrative data is not readily accessible or reusable for analysts, policy advisors, researchers or the general public, nor are the datasets of the various funding agencies linked.

This domain plan seeks to redress this situation over a period of at least five years. The plan outlines a set of staged, coordinated actions across government and the research community. Some actions can be implemented relatively quickly, others will take several years. These actions have been developed following extensive engagement with key government agencies over the past year, and meetings and workshops with the research sector and end users.

The actions cover funding and expenditure, R&D outputs, people and skills, business R&D and innovation, collaboration, knowledge exchange and commercialisation, infrastructure and costs. The emphasis is on improving the quality and standardisation of basic data on the inputs and outputs of the system, and establishing links to the rich data sources of Statistics New Zealand and other government agencies.

This domain plan represents a commitment from government agencies to improve coordination of data and information and to lay the framework for developing a system-wide data infrastructure. Implementation of the action plan should over time lead to the development of a comprehensive national research information system that generates quality longitudinal data and statistics. In line with the Government's data and information management principles, we expect much of the data to be made available and accessible to the public. This will be an invaluable resource for government, the research community and end users in New Zealand.

David Smol, Chief Executive,
MINISTRY OF BUSINESS, INNOVATION AND EMPLOYMENT

Mark Sowden, Deputy Government Statistician and Deputy Chief Executive Customer Strategy and Delivery,
STATISTICS NEW ZEALAND

Katrina Casey, Acting Secretary for Education,
MINISTRY OF EDUCATION

Tim Fowler, Chief Executive,
TERTIARY EDUCATION COMMISSION

Table of Contents

Organisational Endorsements 2

Summary 3

1 Purpose of the domain plan 5

2 Scope of the plan 8

3 Conceptual approach 10

4 Process of developing the plan 14

5 The enduring questions for science and innovation 17

6 Data sources to inform questions 24

7 Enduring data needs 26

- 7.1 R&D funding and expenditure 27
- 7.2 R&D outputs 28
- 7.3 People and skills 29
- 7.4 Business R&D and innovation 30
- 7.5 Collaboration 30
- 7.6 Knowledge exchange and commercialisation 31
- 7.7 Infrastructure 32
- 7.8 Costs 32

8 Analysis 33

- 8.1 R&D funding and expenditure – Who's doing what research? Who's funding the research? 33
- 8.2 R&D outputs – What's being produced? 36
- 8.3 People and skills 37
- 8.4 Business R&D and innovation 38
- 8.5 Other gaps 38

9 Set of actions 40

10 Implications of a national research information system 45

Glossary of terms 47

Appendix A: Defining and categorising R&D 48

Appendix B: Details of current data holdings 51

Appendix C: Data fit-for-purpose assessment 58

Appendix D: Links between enduring questions and data questions 69

Appendix E: ORCID identifiers 78

References 80

Organisational Endorsements

Callaghan Innovation

Callaghan Innovation is pleased to endorse the development of the Research Science and Innovation Domain Plan report. As a funding agency of R&D Grants, this plan will enable us, through better access to information, to improve and have a greater understanding of the evaluation and impacts from government investment in research, science and innovation. We look forward to having access to more accurate, reliable, and timely data to help us more effectively support businesses across the innovation sector.

Health Research Council of New Zealand

The Health Research Council of New Zealand (HRC) welcomes the opportunity to formally endorse the *Research, Science and Innovation Domain Plan*. The completion of this plan demonstrates the shared commitment of government agencies to developing a system-wide data infrastructure that should improve our collective ability to generate timely, quality and reliable data on the science and innovation system. The actions set out within the domain plan, once achieved, will deliver an important resource for all those involved in the science system. The HRC anticipates it will provide us with a broader, system-wide view of New Zealand's health research investment which will support us in developing evidence-based research investment strategies to improve the health and wellbeing of New Zealanders and support stronger economic growth for New Zealand.

Ministry for the Environment

The Ministry for the Environment strongly supports the intent of the domain plan to improve national information and data on our science system. We look forward to the implementation of the proposed action plan.

Ministry for Primary Industries

The Ministry for Primary Industries (MPI) is strongly supportive of the intent of the *Research, Science and Innovation Domain Plan* and actions outlined within it. Successful implementation of the Plan will enable better coordination of research, science and innovation data, thereby providing a solid base for: assisting MPI in its assessment of the impact of our investment in research, science, and innovation; setting policy priorities for science and innovation as they relate to our sectors; and improving reporting, accountability and transparency across government. Ensuring flexibility of the system-wide data and information standards that arise from the Plan will be essential in capturing the breadth of activities across the primary industries and wider science and innovation system. MPI is committed to working with other funding agencies on successful implementation of the Plan.

Royal Society of New Zealand

The *Research, Science and Innovation Domain Plan* is a practical extension of the Government's National Statement of Science Investment, which promises to lead to improvements in the transparency and discoverability of New Zealand research, while decreasing transaction costs across the system. Evaluating the effectiveness of research policies and mechanisms is particularly important for a small nation like New Zealand to ensure it applies its limited resources wisely. So the Royal Society of New Zealand supports the intent of the Domain Plan in the expectation that it will significantly improve the coverage and cohesion of relevant data on the research system and lead to more evidence-based decision making around research investment and government policy.

Summary

Information and data on New Zealand's innovation system has suffered from a lack of oversight and coordination for many years. This has led to problems with data integrity, data structure and standardisation, reporting capability and data validity. Much of government agencies' administrative data on research and science investment is not readily accessible and reusable by analysts, policy advisors, researchers or the general public, nor are the datasets of the various funding agencies linked. As the funding landscape has become more complex, with additional funding mechanisms and increased devolvement of decision making, the problem has become more acute. At the same time, increased investment in the sector has intensified the pressure to demonstrate results and impacts, lift transparency and reduce transaction costs.

The domain plan covers data on New Zealand's innovation system.[1] The plan covers all areas of research and development (R&D), which includes the natural and social sciences, engineering, mathematics, technology, the arts and humanities. The purpose of the domain plan is to achieve clarity and broad agreement on the main statistical and information priorities for research, science and innovation in New Zealand. In response to these priorities, the domain plan also sets out a programme of coordinated actions across government and the research community to respond to the identified priorities. The actions seek to strike a balance between addressing the diverse needs identified and the cost of data collection.

The domain plan does not seek to determine performance or impact indicators, nor does it seek to generate an evaluation framework. Rather, the intention is to create an underlying data infrastructure that can be used for multiple purposes, one of which is to generate performance measures for New Zealand's research, science and innovation system.

The Ministry of Business, Innovation and Employment (MBIE) has led the domain plan process, engaging extensively with the key government agencies in the innovation system over the past year, and facilitating meetings and workshops with the research sector and business community. The process has revealed significant interest in improving data and its collection. Funding agencies have been principally interested in better understanding the system and the results of their investments. The research community has shown particular interest in understanding the research landscape, streamlining data collection frameworks and making use of smart information systems.

Participants in the domain plan have identified the strategic knowledge needs for research, science and innovation and formulated them into six broad questions. These are called enduring questions and cover impacts, return on investment, design of investment and the role of people. These questions ultimately drive the data needs, which have been categorised into eight topics: funding and expenditure, R&D outputs, people and skills, business R&D and innovation, collaboration, knowledge exchange and commercialisation, infrastructure and costs. The domain plan seeks to improve our ability to answer these questions over time through improvements to science and innovation data.

1 The focus of the plan is on administrative data and the meta-data associated with research outputs, not on all the data that researchers create and use in the course of their research.

Analysis of current data holdings and the links to the enduring questions identified that the following areas most need addressing:

1. research being performed in New Zealand
2. funding sources of research, particularly government sources
3. research outputs – quantity, quality and use
4. research capacities and capabilities, including research mobility and the absorptive capacity of business
5. research collaboration and links with end users.

Data on business R&D, firm innovation and skills for innovation is in general relatively fit-for-purpose. However, there are significant data gaps on the inputs and outputs of research. Without this data, impact and evaluation analysis will remain difficult and transparency will be limited. High-level outcome measures, such as industry-level productivity, disease burdens, or biodiversity, already exist across different government agencies in New Zealand and these need not be the focus of the domain plan. Reliable input and output data on the research and science system forms the basis for attribution analysis to these higher level outcomes.

Addressing the key data gaps identified will require coordinated action across government and the research community. This report identifies a set of 26 actions, which seek to improve data sharing and integration, access to data, data quality, reusability and the ability to aggregate across the system. The actions also seek to reduce collection burden through consistent system-wide data standards, connecting systems and the use of modern information systems to collect data.

Aggregation of government funding data will require a set of standards be developed and adopted across the system. With the exception of the Performance-Based Research Fund, this will require data to be collected at a project level across all funds, with consistent data collected on the types of research undertaken, the research teams, end users, collaborations, sub-contracting and research outputs generated. This report also foresees ORCID[2] identifiers for researchers being used and adopted across the research system, which will be critical in unambiguously linking researchers, their outputs and funding sources.

These actions will form the basis of a national research information system, something many other countries already have in various forms. Developing such a system is essential for the efficiency and competitiveness of New Zealand's research sector. The New Zealand data system will be linked to Statistics New Zealand Longitudinal Business Database and the Integrated Data Infrastructure. In line with the Government's data and information management principles, the data will be made available and discoverable to the public to the greatest extent possible. This will be an invaluable resource for the research community, end users and government agencies. It will also raise the visibility and profile of New Zealand's system internationally.

The actions outlined in this report also include a revision to the R&D Survey and further minor improvements to the innovation module of the Business Operations Survey. Establishing a method for deriving a reliable annual figure for business expenditure on R&D is a separate action that will receive high priority. Other actions relate to improving data on commercialisation, infrastructure and costs.

---

2 ORCID stands for *Open Researcher and Contributor Identifier.*

1 Purpose of the domain plan

The purpose of the Research, Science and Innovation Data Domain Plan is to achieve clarity and broad agreement on the main statistical and information priorities for research, science and innovation, and to provide the strategy and key actions for addressing these priorities over the coming years. The domain plan puts forward:

- A long-term picture of what is required to improve official statistics, data and information on research, science and innovation in New Zealand
- A coordinated plan for addressing issues
- A cross-agency approach to long-term statistical, data and information priorities.

The National Statement of Science Investment contains a vision for 2025 of "a highly dynamic science system that enriches New Zealand, making a more visible, measurable contribution to our productivity and wellbeing through excellent science". This includes a "comprehensive evaluation and monitoring of performance, underpinned by easily available, reliable data on the science system, to measure our progress towards these goals". The domain plan contributes to that vision by establishing a framework for a system-wide data infrastructure. The domain plan also supports the Tertiary Education Strategy, which specifies strengthening research-based institutions as a priority. The strategy also contains the following indicator of success: tertiary education organisations develop strategies and monitoring systems to measure their progress in contributing to innovative activity.

The domain plan will inform changes to the collection, analysis and dissemination of data. It aims to ensure that the data being collected is relevant, useful and meets future needs. The collection of statistics and data incurs certain costs, particularly on researchers and their institutions. The domain plan seeks to strike a balance between responding to the key data needs and the cost of collection. It is important to note that research information systems which make use of information and communications technology (ICT) systems are significantly lowering the cost of data collection and improving data quality. Use of these systems, adoption of data standards across the system and linking datasets is expected to reduce transaction costs significantly.

As with other domain plans, this plan addresses information needs in a coordinated and collaborative way. This is particularly important for this plan because of the diverse needs and the wide range of actors and agencies with a stake in the innovation system. The users of the information and data are many and span different levels and parts of the system. The key users include:

- Parliament and Ministers
- policy agencies
- funding agencies
- the research community
- end users of research, including business, industry organisations and local government entities
- the general public.

Each user group has different, but often overlapping, needs. At a broad level they are:

- **Parliament and Ministers** require information for accountability purposes and for taking decisions on orienting the system and designing innovation policies
- **Policy agencies** require information to provide advice on system design, policy priorities and funding allocations, and to undertake evaluations
- **Funding agencies** require information on where and how to invest funding and on what research other agencies are funding

- **The research community** requires information on research priorities, the research landscape and the needs of end users
- **Business** requires information on funding sources, where research expertise lies and with whom to partner
- **End users of research** require information on strengths in the research community, what research is being undertaken and the results
- **The general public** requires information on how taxpayer money has been spent and the results of that spending.

What will success look like?

More accurate, reliable, accessible and timely data

Government agencies and funders

- A **greater understanding** of how New Zealand's innovation system works and performs
- More **evidence-based policy settings** in research, science and innovation
- Better **evaluation of the impact** of government's investment in research, science and innovation
- **Smarter investment decisions** in research, science and innovation
- **Enhanced design** of business support for innovation and R&D
- **Improved transparency** of the research and science system

Researchers, firms and New Zealand public

- **Greater public accessibility** of research, science and innovation data
- **Reduced transaction costs** for researchers in applying for funding and in reporting
- **More efficient and effective processing** of research grant applications and reporting data
- **Greater visibility** of New Zealand's innovation system, both domestically and internationally

The needs of policy and funding agencies are perhaps the largest given their roles in the system, although it is important to note that many of their information needs are similar to those of the research community. Aside from their day-to-day operations, policy and funding agencies are frequently asked questions of the following nature:

- What is the effectiveness of a particular mechanism or fund?
- How does our innovation system perform relative to other countries?
- Which industries are most active in R&D?
- How does our firms' investment in R&D compare with other countries?
- Does New Zealand need researchers in a particular area?
- What is the basis for funding decisions?
- How much funding was provided for a particular topic, such as renewable energy or soil science?
- In what cities and regions is research taking place?
- Is a certain researcher involved in any contracts?

- Where are international collaborations occurring?
- With which institutions overseas do our researchers collaborate?
- What capability does New Zealand have in a particular field?
- How many researchers have migrated to New Zealand from other countries?
- How much funding is directed to Māori researchers and research of particular relevance to Māori?
- How many female researchers received funding?
- How many postdoctoral researchers does the government support?

As connections deepen across national innovation systems, it is important that New Zealand data is internationally comparable as far as possible. This will assist collaboration between businesses and researchers across borders and provide a better basis for policy advice by enabling cross-country comparisons.

2 Scope of the plan

Innovation comes from the activities of many diverse actors, businesses, multinational enterprises and start-ups, as well as public research institutes and universities. These actors cooperate and compete with each other. Government plays a crucial role in dynamising and orienting this system, by influencing framework conditions, setting innovation policies and providing funding assistance.

The scope of science, technology and innovation policy has progressively widened over time. This follows changes in the conceptual approach to innovation, whereby the innovation system is no longer conceived of simply as the research and science system, but as encompassing skills, tax, procurement, enterprise policy, framework conditions such as market regulation and competition, and environmental policies.

This broadening in the conceptualisation of the innovation system presents a challenge in defining and managing the scope of this domain plan. We made a practical decision on the scope of this plan based on a scan of the quality of data across the main areas influencing innovation policy and on what is feasible to address in a coordinated fashion. Note that the domain plan is concerned with data and statistics on the research system itself, i.e. administrative data. The data that researchers create and use in the course of their research is not the focus of this project, except for relevant administrative data such as how 'created data' is used. The domain plan covers all fields of research and science, measures of business innovation and public support for innovation. It also includes all activities of the public science system, such as testing and standardisation, and general-purpose data collection and documentation which may or may not be part of the R&D process.

What is in and out of scope?

In scope

- The R&D system
- Business innovation
- Commercialisation
- Skills pipeline for R&D
- Management skills and entrepreneurship
- Coordination and connectivity within the R&D system and between the R&D system and users

Out of scope

- Macroeconomic conditions
- Regulatory and IP settings
- Competitive environment
- Enterprise policy
- Access to capital
- Generic skills
- Industry and firm structure

The areas that are out of scope are critical influencers of the innovation system, but data for most of these areas is already of considerable quality and coverage. The skills area is the most challenging scope issue. In higher education institutions, research and teaching are closely linked as most academic staff undertake both. The results of research feed into teaching, and the information and experience gained in teaching can often result in an input to research. It can therefore be difficult to determine where the education and training activities of higher education staff and their students end and R&D activities begin, and vice versa.[3] It is important therefore to recognise the

3 This issue also affects the measurement of health-related R&D. University or teaching hospitals train medical students in addition to their primary activity of healthcare. The activities of teaching, R&D, advanced medical care and routine medical care are frequently closely linked. See OECD (2001).

link between research and teaching and to recognise that the tertiary system trains people, many of whom leave the R&D system. This contribution to human capital is an important underpinning aspect of the innovation system. For practical reasons, data on generic skills is excluded from this domain plan, but skills specifically for the R&D system are included. Data on researchers, which includes PhD students and some Master's students, is in scope. The framework diagram in the following section provides further clarification.

It is important that New Zealand data on innovation and R&D is comparable internationally for policymaking purposes. We are therefore guided by the key international manuals in this area, i.e. the Oslo Manual for innovation and the Frascati Manual for R&D. See Appendix A page 48 for further details on defining and categorising innovation and R&D.

3 Conceptual approach

Policymakers and innovation researchers use two broad approaches when thinking about the innovation system.

1. The first approach is the *linear model of innovation*, which emphasises a science push model. Under this model, researchers receive funding and generate new knowledge and ideas which are then picked up by industry and other end users. The linear model is associated with a traditional input-output model.
2. The second approach, which has grown in popularity over the last 20 years since the OECD's 1997 publication, is the *systems approach* to innovation.[4] This approach emphasises the connections and feedback loops in the system, such as those between researchers and end users where end users may influence the research agenda. The systems approach emphasises demand-pull, co-innovation and collaboration. Over the past decade the systems approach has brought into the frame the effect of other policies on innovation, such as broader regulatory settings, competition and access to capital.

The framework for the domain plan is guided by both approaches, as we consider that neither model fully captures how the system works. The framework diagram on the following page shows the inputs, outputs and outcomes of the innovation system, which the linear model emphasises. It also shows the key connections and feedback loops in the system, which the systems model emphasises.

The framework diagram also shows what is in and out of scope of the domain plan. Areas with no information or data details are excluded from the domain plan. The details in the boxes show desirable key information that is needed to deliver on the objectives of the domain plan. They include information on research outputs and knowledge exchange or engagement activities.

The framework shows that research, science and innovation contribute to multiple outcomes: economic, social, health and environment. Vision Mātauranga objectives are high-level outcomes that contribute to all outcome areas. These outcomes align with those presented in the National Statement of Science Investments. Research and science contributions to public policy are often very significant, but they are an intermediate outcome rather than a final outcome, i.e. better public policy is not an end goal in itself.

The framework guides the choice of data collected and the interpretation placed on any indicators that can be constructed from the data. As with any framework, it is an abstract representation of the real world and is designed to capture the underlying complex reality in a simplified manner. It helps organise ideas, drive what data should be collected and supports interpretations of the data. At the same time, researchers and others will use the data to test various models which may suggest changes to the framework and data needs over time.

The domain plan is also guided by the Government's broader approach to data. The Open Government Information and Data Programme emphasises the availability and discoverability of data. In line with the programme's aim and the New Zealand Data and Information Management Principles,[5] the domain plan is underpinned by the following principles:

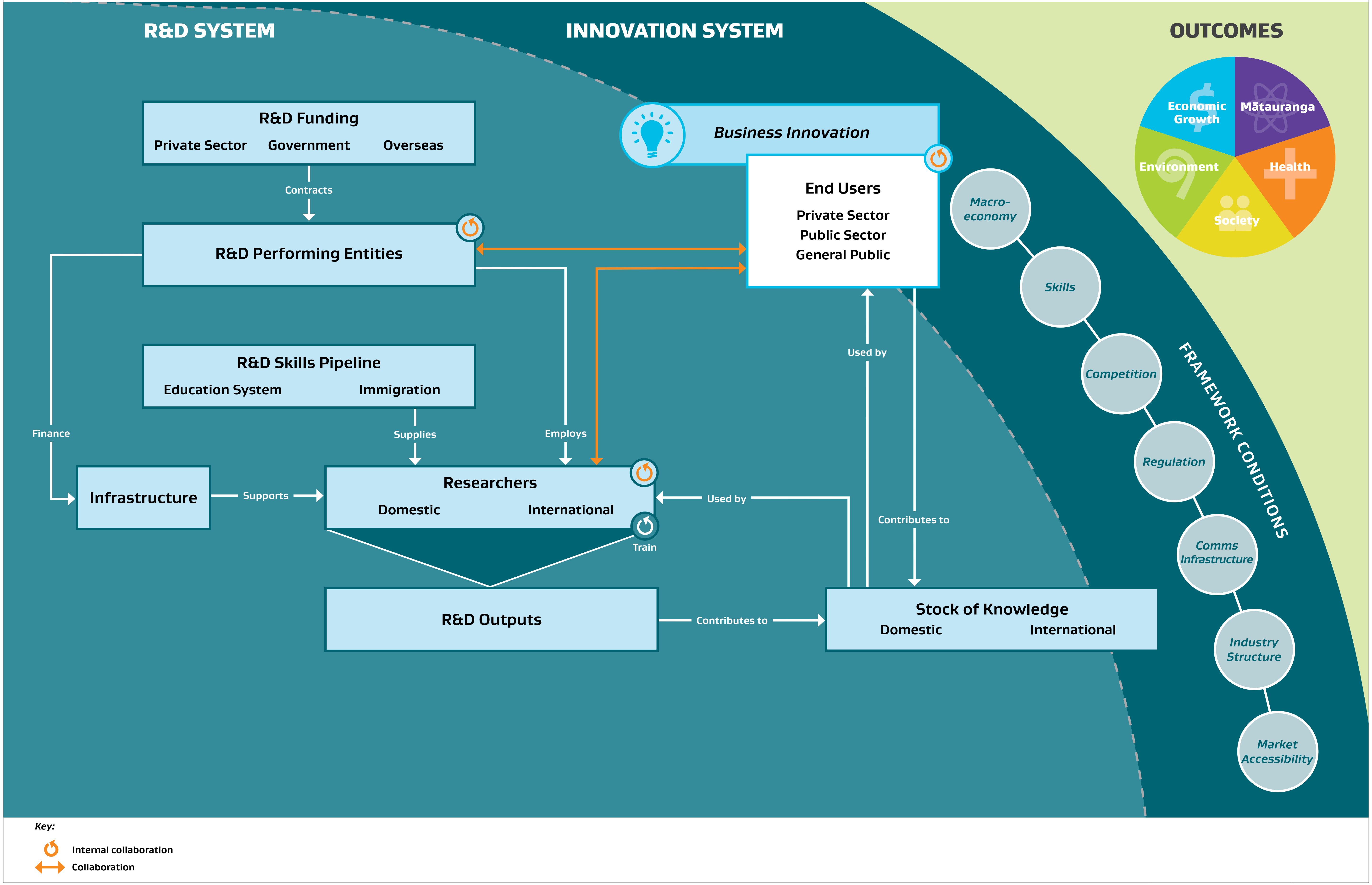
1. Provide a **system-wide view** of science and innovation
2. Ensure **open** data, **accessible** easily and widely
3. **Protect** personal and commercially sensitive data
4. Enable the **reuse** of data
5. Reduce **collection burden**
6. Ensure data is **trusted, authoritative** and **well-managed**
7. Automate **system connectivity**.

4 See Lundvall (1992), Nelson (1993) and OECD (1997).

5 CAB Min [11] 29/12.

MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

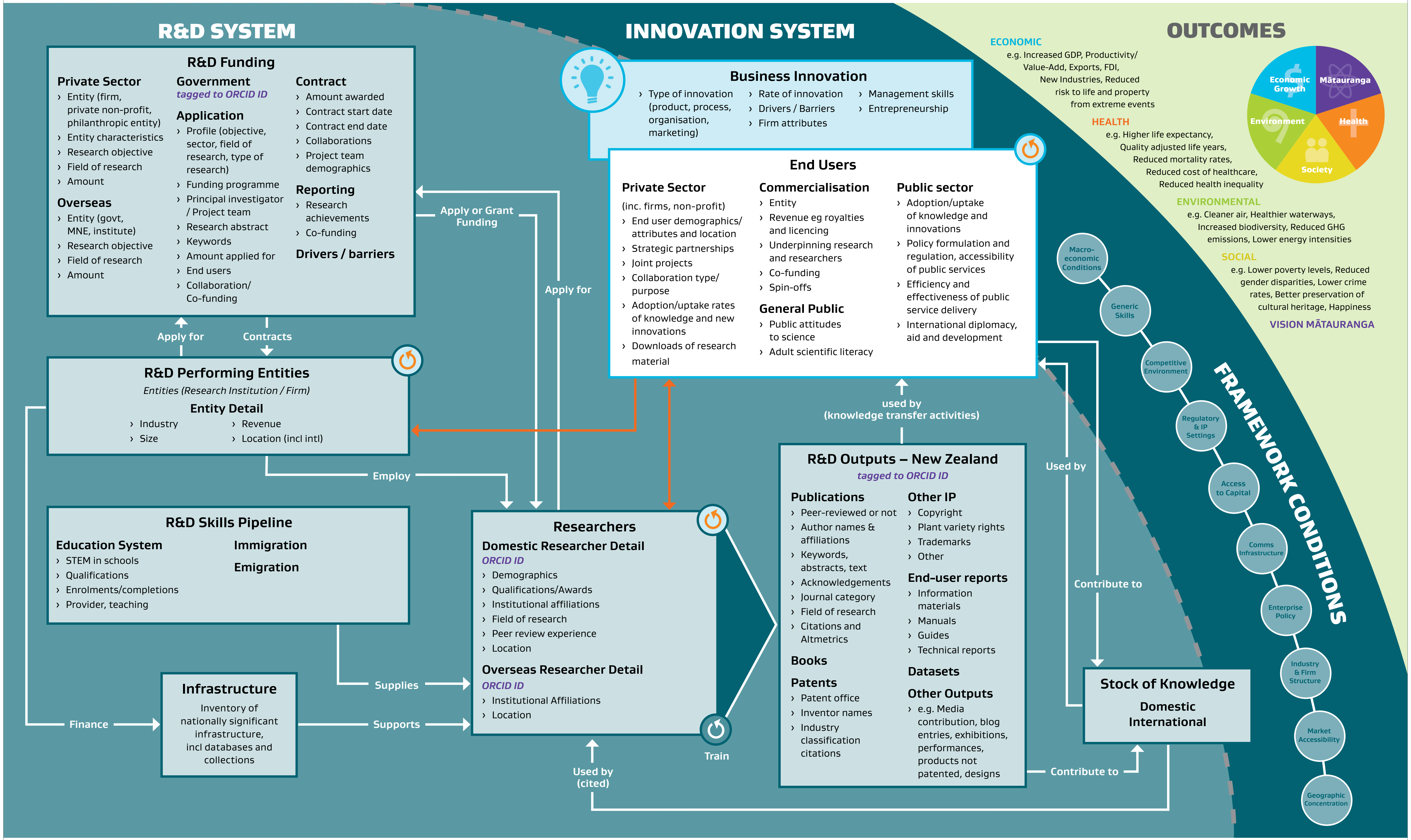
NZ Science and Innovation System



Research, Science and Innovation Domain Plan Framework

Guiding Principles

1. Provide a system-wide view of science and innovation
2. Ensure open data, accessible easily and widely
3. Protect personal and commercially sensitive data
4. Enable the reuse of data
5. Reduce collection burden
6. Ensure data is trusted, authoritative and well-managed
7. Automate system connectivity



Key

- Internal Collaboration
- Collaboration & Knowledge Flow

Participants

1. Ministry of Business, Innovation & Employment
2. Ministry of Education
3. Tertiary Education Commission
4. Statistics New Zealand
5. Ministry for Primary Industries
6. The Treasury
7. Royal Society of New Zealand
8. Health Research Council
9. Callaghan Innovation

Input received from the research community, including Universities, CRIs and independent research organisations

MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

4 Process of developing the plan

The development of the domain plan took place over a 12-month period. MBIE led the process,[6] working particularly closely with Statistics New Zealand, the Ministry of Education, the Tertiary Education Commission and the Ministry for Primary Industries.

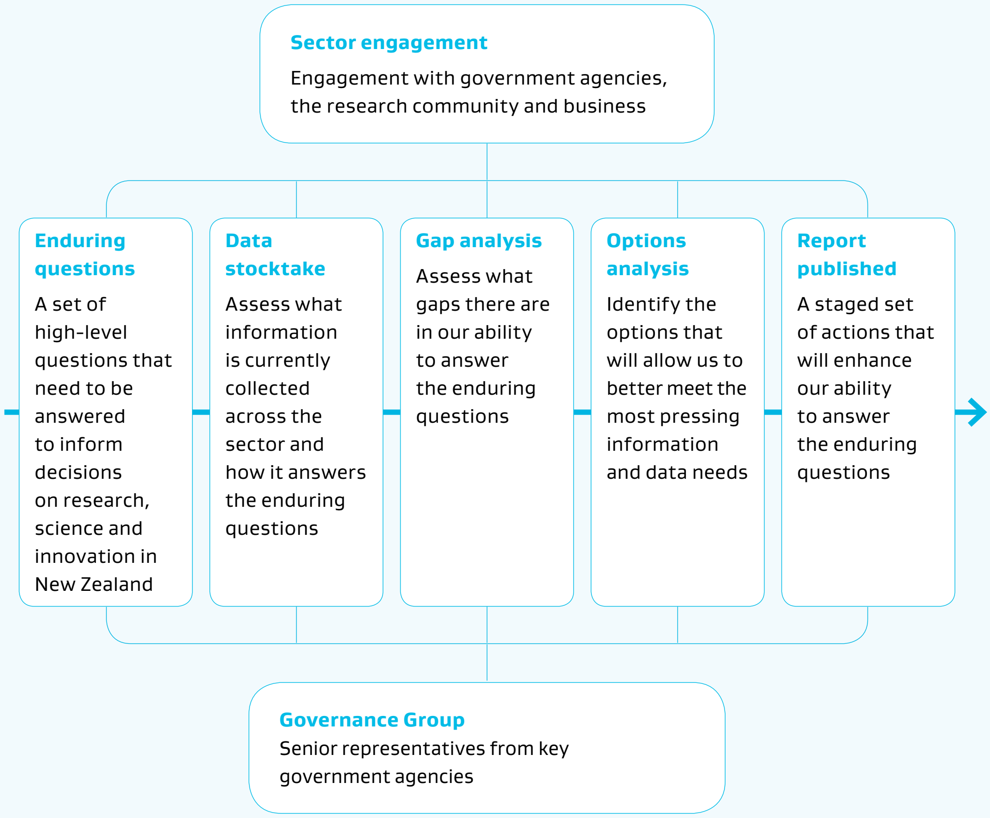
Because of the importance of research and innovation to a number of government agencies, an inter-agency governance group oversaw the domain plan process.[7] The governance group will also oversee the implementation of the domain plan work programme which is expected to take several years.

As with other domain plans, the process involved identification of enduring questions, a data stocktake, gap analysis and development of actions (see diagram below). As part of the process we have examined international developments in research information systems and impact assessment. This includes euroCRIS (Current Research Information Systems), Flanders' FRIS, the Netherland's National Academic Research and Collaborations Information System (NARCIS), Norway's CRIStin, Portugal's PTCRIS, Sweden's SweCRIS, the UK's Gateway to Research and Researchfish, the US Federal RePORTER which is part of the STAR METRICS initiative, and the grants and projects portal of Research Data Australia.

6 The lead role of MBIE reflects the wide-ranging information and data needs of the Ministry. MBIE requires information for the following purposes: (1) providing policy advice on the innovation system; (2) evaluation of research, science and innovation policies and programmes; (3) contract management of MBIE's science investments; (4) monitoring of Crown Research Institutes; (5) monitoring of funding administered by other agencies through Vote Business, Science and Innovation.

7 The governance group comprises senior managers from MBIE, Statistics New Zealand, the Ministry of Education, the Tertiary Education Commission, the Ministry for Primary Industries, Callaghan Innovation, the Treasury, and a member of the MBIE Science Board.

Domain Plan Process



Due to the importance of data coordination across government and funding agencies, the domain plan process began with a series of workshops with government and funding agencies.[8] This was followed by engagement with the research community[9] and business[10] on knowledge needs, data gaps and solutions.

Following these two processes, MBIE convened a working group across government agencies and the research community to discuss key data gaps and solutions. It is expected that the working group will be critical to implementing the work programme contained in this report. MBIE also convened a working group on ORCID[11], with representatives from funders and the research community.

The following timeline shows the domain plan process.

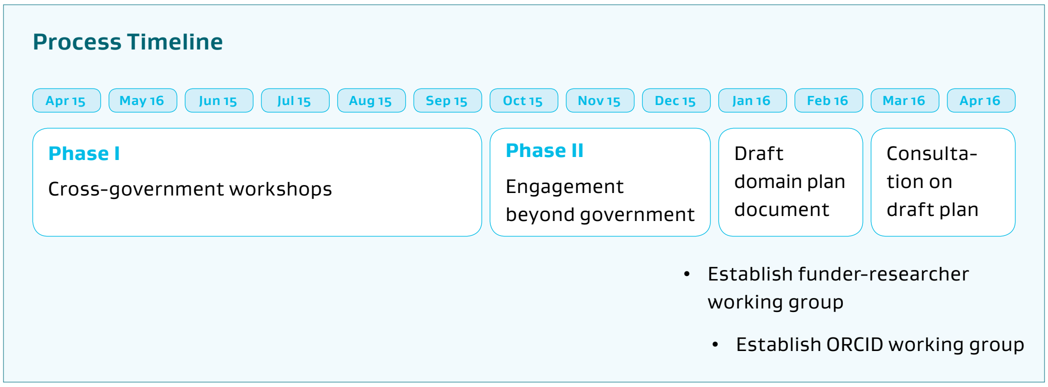
---

8 MBIE, the Ministry of Education, the Tertiary Education Commission, Statistics New Zealand, the Ministry for Primary Industries, Callaghan Innovation, the Treasury, the Health Research Council, the Royal Society of New Zealand, and the Ministry for the Environment.

9 Engagement with the research community involved discussions with University Deputy Vice Chancellors, University research office staff, University librarians, some faculty staff members, several University technology transfer offices, CRI strategy members, CRI impact, planning and evaluation managers, the Independent Research Association of New Zealand, and the Association of the Centres of Research Excellence.

10 Engagement with business was more limited than that with the research community due to the greater implications for the research community. Meetings were held with Business NZ and several industry organisations.

11 ORCID stands for *Open Researcher and Contributor Identifier*. See Appendix E for further information.



In addition to researchers and several business groups, MBIE discussed the domain plan and received advice from the MBIE Science Board, Departmental Science Advisors and the economic development agencies of New Zealand. MBIE met several times with the Ministry for the Environment, the Ministry of Health and the Productivity Commission to provide updates on the domain plan.

5 The enduring questions for science and innovation

This domain plan examines the research, science and innovation needs of Parliament and Ministers, policy advisors, funding agencies, research providers, business, end users of research and the general public. The domain plan addresses the information required for policy and evaluation purposes, for accountability purposes arising from the use of public money, and for decisions by the various actors in the system, including Universities and research organisations.

Basic input and output data needs to be reliable, accurate, timely and as complete as practicable. Consistency in data parameters and collection across actors, in particular funding agencies, is essential for generating system-wide data. Indicators of outcomes and impacts are also important, but the domain plan does not seek to determine performance indicators, nor does it seek to generate an evaluation framework.

Certain questions will endure for research, science and innovation over the long term. While we may not necessarily be able to answer these questions very well, we can identify them and put in place the systems needed to ensure that data and statistics are increasingly available so that they can better be answered.

Domain plans use the enduring question construct to frame strategic knowledge needs. These questions help categorise and structure the things that 'we know we will need to know' into the future to inform policy, strategy and decision making. By their very nature as *enduring* questions, they are difficult to answer and it is likely that over time answers will change.

These enduring questions have not been prioritised. The importance of each question is likely to shift over time and it is difficult to reach consensus on their importance. What is certain is that these questions are ultimately the drivers of national-level statistics, data and information. Economic objectives have been separated only because there are specific data sources that apply to measuring economic outcomes. This is not to suggest that economic, social, health and environmental outcomes are discrete and unrelated to each other.

Our Enduring Questions

1. What is the **contribution** of research, science and innovation **to economic objectives**?
2. What is the **contribution** of research, science and innovation **to social, health and environmental objectives**?
3. What is the **social return to the marginal dollar of government investment**?
4. How should government **design its investments in research and science**?
5. How should government **design its support for business R&D and innovation**?
6. What is the r**ole of people** in the research, science and innovation system?

Question 1: What is the contribution of research, science and innovation to economic objectives?

There is widespread agreement among economists and policymakers that new knowledge production and innovation are central to the trend rate of economic growth.[12] In OECD economies, only educational attainment and research intensity have consistently been significant contributors to multifactor productivity over the past 140 years.[13] The importance of innovation is only likely to grow as global competition strengthens and firms are forced to compete on value and product differentiation which is underpinned by innovation. Innovation, particularly process innovation, is also important in reducing cost structures and raising efficiency.

The contribution and value of research, science and innovation to economic objectives is ultimately measured by improvements to productivity at a national level, or industry and firm level. This requires reliable measures of inputs, such as expenditure on R&D, and economic activity measures. Econometric studies can be used to establish the strength of the relationship between economic value-added and research and science inputs. However, this technique is silent on the mechanisms by which innovation contributes to increased productivity, unless the specified model is testing a particular hypothesis such as that distance inhibits technology transfer. Econometric studies therefore need to be supplemented by other studies which draw on research outputs and qualitative and quantitative information on how research outputs lead to outcomes.

The following questions sit under this high-level question:

a. What is the effect on production and productivity (aggregate, regional, industry-level) of R&D performed in the public sector (government and higher education)?
b. What is the effect on firm performance of business expenditure on R&D (BERD) and innovative activity?
c. What is the contribution to productivity of different areas/types of research and of R&D financed by government, higher education, the private sector (including industry levies) and overseas? Which funding decisions result in greater economic impact?
d. To what extent are R&D and innovation driving structural change in the economy (new firms, new industries, new exports)?
e. What is the contribution of research to economic-related public policies?
f. To what extent are science and research driving Māori and Pacific economic development?
g. To what extent are science and research responding to regional economic development needs?

Question 2: What is the contribution of research, science and innovation to social, health and environmental objectives?

Research, science and innovation can make significant contributions to the achievement of social, health and environmental objectives. These objectives can be ends in themselves, but they can also contribute to the achievement of economic objectives. For instance, a healthier population can reduce the financial burden of ill-health and can lead to higher levels of labour productivity. Likewise economic outcomes may feed into social and environmental outcomes. For instance, higher incomes may lead to more investment in the health sector or to more funding available for energy-efficient technologies.

Advances in health research can lead to the eradication of diseases, reduction in burden of disease, improvements in longevity and quality of life, better nutrition and more effective and efficient delivery of healthcare. Other areas of social research can improve the quality of life of children and families, education, housing, and public order and safety. Research may also lead to a better understanding of culture and identity which can improve wellbeing.

---

12 OECD (2010).

13 Madsen (2001).

Research may also lead to better long-term management of our environment and natural resources and to fewer threats to biodiversity. Better understanding of the natural environment, generated through research, allows more informed decision making and can lead to better environmental, economic and social outcomes for New Zealand. This is particularly important given current and emerging environmental challenges including climate change, biodiversity loss and freshwater management. Researchers may also develop technological solutions that can contribute to predicting, monitoring and responding to environmental challenges, or create more environmentally friendly products and processes. Environmental research may also lead to better public policy settings which reduce the risk to life and property from extreme events.

Answering this question requires datasets of social, health and environmental outcomes. The Longitudinal Business Database can be used for economic impact assessment, but cannot be used for these non-economic outcomes, hence their separation from the economic outcomes in the previous question. Data holdings of the Ministry for the Environment, the Ministry for Social Development, the Ministry of Health and Statistics New Zealand are relevant to this question.

The following questions sit under this high-level question:

a. To what extent does the research and science system generate research that is useful for addressing social and health issues of particular relevance to New Zealand, including its demography?
b. To what extent is the research and science system contributing to the Vision Mātauranga policy and how? How is the Vision Mātauranga policy contributing to the research and science system?
c. How does the research and science system perform with respect to inclusiveness, such as gender equality or ethnic representation?
d. To what extent does the research and science system generate research and data that improve the understanding of New Zealand's unique natural environment and assist with environmental monitoring and reporting?
e. How does research support sustainable environmental management, including the sustainable use of New Zealand's natural capital and resources?
f. To what extent does the research and science system generate knowledge that assists with reducing risk from extreme events?
g. To what extent does the research and science system contribute to international diplomacy and contribute to aid and development?
h. What is the contribution of research to public policies in the social, health and environmental areas?
i. What value do New Zealanders place on non-economic outcomes?

Question 3: What is the social return to the marginal dollar of government investment?

The primary rationale for government support of R&D is that firms underinvest in R&D because of uncertainties and the lack of appropriability, ie firms cannot capture all of the benefits from investment in R&D. Empirical studies fairly consistently support the theory, finding that the private rates of return to R&D are high and that the social returns to R&D are higher than the private returns.[14] However, there is little consensus on which industries generate higher returns and by how much these returns exceed those of other industries. The justification for government support is strongest in basic research, but there is little evidence on what types of research generate higher returns.

14 Empirical studies (mostly from the USA, Japan and France) find that the estimated private returns from R&D range from 7% to 69% (see Wieser 2005).

International literature suggests that publicly-funded research increases the stock of knowledge, creates human capital and builds absorptive capacity, facilitates and improves private sector R&D, leads to innovation, develops networks and stimulates social interaction.[15] Scientific data supported by government can also underpin a wide range of operational decision making and policy formation. Studies assessing the effects of public R&D are less conclusive than those assessing private sector R&D, but often find positive effects.[16] With the exception of the USA, the majority of studies find a degree of complementarity between public and private sector research activity.

There are few New Zealand studies on the impact of R&D at a sector or national level.[17] Most studies tend to be qualitative, adopt case study approaches and focus on particular programmes.[18] It is therefore at present difficult to draw conclusions on the effects of New Zealand R&D – whether private or public – on economic growth or on environmental and social outcomes. However, the demonstrated importance of research and innovation at an international level would suggest that government has an enduring role to play in orienting and funding the system in part.

According to the National Statement of Science Investments (NSSI), the government currently invests around $1.55 billion annually in research and science. The government is the shareholder of the Crown Research Institutes, the primary funder of the higher education sector and offers support for business R&D and innovation. Funding for R&D has been progressively rising over recent years.

The following questions sit under this high-level question:

a. What are the impacts and spillovers of government-funded R&D for different types of research and funding mechanisms? Given this, what is the appropriate co-funding rate?
b. What is the extent and nature of intra-industry and inter-industry R&D spillovers? Given this, what is the appropriate co-funding rate?
c. What is the relationship between government funding for R&D and private sector funding for R&D?
d. What contribution does foreign R&D make to New Zealand?

Question 4: How should government design its investments in research and science?

As outlined above, there is a strong case for why government should invest in knowledge creation and innovation. However, there is little international evidence to draw on with respect to the design of those investments. The innovation literature emphasises that investments should respond to the specific institutional settings and characteristics of a particular country. The application of economic principles to science policy emphasises competition and incentives.

15 OECD (2013) and Salter & Martin (2000).

16 OECD (2005).

17 A study on agricultural R&D from 1927 to 2001 found that domestic knowledge, both private and public, was positively associated with productivity growth, generating an annual return of 17% over the period, but much less so than for foreign knowledge (Hall and Scobie 2006). It is likely that domestic capability was vital in absorbing and using knowledge generated offshore. This finding is consistent with the international literature which finds strong returns from public R&D in the agriculture sector. However, another study of nine industries in New Zealand found publicly provided R&D had no impact on productivity (Johnson, Razzak and Stillman 2007). The industries in that study covered the primary, manufacturing and services sectors. An MED evaluation in 2011 (MED 2011) found that public support of private R&D had a significant positive additional impact when targeted at firms that were building capability, that were small and that had not previously undertaken R&D.

18 See, for instance, NZIER (2015) on the Cawthron Institute.

Each country has a unique innovation system which is formed by its unique history, institutions and policy settings. There is no one optimal shape of a system. That said, high-performing systems share a number of characteristics. They:

i. Develop and maintain strong connections and linkages between research and industry and other end users
ii. Generate innovation within and across all sectors and segments and demographic groups of society
iii. Are driven by a strong business sector that undertakes the major portion of R&D
iv. Allow private sector innovation and R&D to flourish and publicly provided R&D complements that of the private sector rather than displaces it
v. Adapt and respond to changing needs and emerging issues, but at the same time the system as a whole is relatively stable
vi. Produce excellent research with an appropriate balance carried out across the research spectrum (basic, applied, mission, experimental, translational)
vii. Sustain high levels of absorptive capacity and are quick to adopt the latest ideas and knowledge and dissemination happens rapidly
viii. Maintain broad capabilities across research fields as well as depth in particular areas
ix. Show high levels and depth of connections internationally
x. Are underpinned by strong framework conditions for innovation such as competition, open markets and law
xi. Are shaped effectively by an appropriate balance of top-down and bottom-up policies and measures
xii. Remain free from blockages so that information and ideas flow freely through the system, its institutions and actors
xiii. Exhibit high levels of collaboration in the research community.

The following questions sit under this question:

a. How do research capacities in New Zealand improve New Zealand's ability to absorb knowledge and ideas generated offshore?
b. To what extent should New Zealand specialise in its investments in public R&D versus maintain a broad base?
c. To what extent does New Zealand R&D contribute to the global pool of knowledge?
d. How well are universities and CRIs responding to end user needs, including knowledge and technology exchange activities? What end user measures are appropriate?
e. Is research funded by different funding mechanisms overlapping or complementary and are there gaps in what is funded? What is the quality of research funded by particular funding mechanisms?
f. How are R&D outputs being used – the extent, the lag, the barriers?
g. What outcomes are associated with particular research outputs, types of research, funding mechanisms (eg competitive versus collaborative) or contract types?
h. What is the relationship between science quality (eg citation rate) and economic, social, health and environmental impacts?
i. What is the impact of collaboration on the quality and productivity of research? What are the barriers to and drivers for collaboration? Is collaboration being strengthened by policy settings?
j. What is the appropriate balance of investment in knowledge creation versus other activities that lead to value (eg start-up support or knowledge exchange)?

k. What are the results from international connections and collaborations?
l. What is the effect of duplication in funding programmes and double-dipping?
m. What value is generated from investment in physical infrastructure in research and science? Which types of research infrastructure investments yield the highest returns? Which investment mechanisms are the most appropriate?
n. What are funding success rates for each fund by researcher characteristic and institution, and is there any bias?
o. What are the effects of different institutional arrangements?
p. What is the best use of research infrastructure, including facilities? How could they be shared or co-owned?
q. What should be the distribution of costs for government activities supported by research (eg, biosecurity testing, costs, fisheries, cost recovery)?

Question 5: How should government design its support for business R&D and innovation?

There is strong theoretical and empirical evidence linking business R&D, innovation and productivity growth. There is also strong evidence of spillovers from business R&D.[19]

The level of private R&D is not beyond the influence of policy. There is good evidence that both tax credits and grants increase R&D expenditure, with R&D tax credits having a stronger but complementary effect to grants.[20] One recent New Zealand study showed that R&D grants increase the likelihood of filing patents and introducing new goods and services. In particular, receiving an R&D grant almost doubled the probability that a firm introduced goods and services that were new to the world.[21]

There is more mixed evidence about whether government support for R&D contributes to increases in long-term productivity.[22] One meta-study found that the success of R&D subsidies depends heavily upon the method of implementation and policy stability.[23] Productivity improvements in agriculture over the last century in the United States have been achieved largely through federal and state investments in agricultural R&D, but with long time lags – average 24 years, with effects continuing for up to 50 years.[24]

The lack of international consensus and the paucity of New Zealand studies suggests that we need a strong data infrastructure in New Zealand to best answer this question.

The following questions sit under this high-level question:

a. What factors and firm characteristics influence the returns to R&D and innovation, and the absorptive capacity, of New Zealand firms?
b. What are the barriers to firm-driven R&D and how can they be overcome?
c. What difference do R&D grants and direct government support for firm-level innovation make to firms?
d. What are the barriers to knowledge exchange between public research organisations and firms?
e. What particular innovative activities lead to better firm performance?
f. To what extent do innovative practices diffuse across firms and how do they diffuse?
g. How do clusters, hubs, precincts and incubators affect firm performance?

---

19 Davis (2006).

20 Elder et al (2013), Westmore (2013).

21 Jaffe & Le (2015).

22 Pells (2014).

23 Westmore (2013).

24 Alston et al. (2010).

Question 6: What is the role of people in the research, science and innovation system?

For any production function, capital and labour are the two primary inputs. In the innovation system, the role of labour (capabilities and skills) is particularly significant given the importance of ideas and knowledge. Any person can plant the seeds of innovation or can come up with an innovative idea, but knowledge generation and innovation is the essence of the R&D system. The exact role people play in the research and science system, as opposed to capital infrastructure is not known. What types of researchers, technicians and support staff and their mix also requires further investigation.

The following questions sit under this high-level question:

a. What is the relationship between teaching and research? How important is research to improving the quality of teaching and what is the flow-on effect to the productivity of tertiary graduates?
b. What are the strengths and gaps in research capability in New Zealand and do the gaps matter?
c. What capability profile do we need across research providers and end users both now and in the future?
d. What are the issues on which central and local government will need science-based input for developing robust policy, and do New Zealand science providers have the capabilities to deliver this?
e. What is the most effective mix of researchers (researcher attributes, skills, fields of research, etc)? Do ‘stars’ (and their networks) generate higher returns than other types of researchers?
f. What is the relationship between workforce stability and R&D outputs and knowledge exchange?
g. How well do we develop, retain, and attract top scientific, research and innovation talent, including management skills and entrepreneurship?
h. What is the role of immigration in the research, science and innovation system?
i. What are the effects of public engagement with science?

6 Data sources to inform questions

Answers to the enduring questions will be informed by theory, international literature and judgment. Critically, evidence is also expected to underpin policy and investment decisions. New Zealand-specific studies and data on the New Zealand system are thus critical, increasingly so as government seeks to base policies on solid evidence.

Two basic families of science and technology statistics are directly relevant to the measurement of innovation: (i) resources devoted to R&D; and (ii) patent statistics. In the past two decades, data from innovation surveys has complemented these two families. Bibliometric datasets on peer-reviewed literature have been important to measuring the output of research and science, although they have not been used for official statistical purposes. Several other datasets are also important to answering science and technology questions:

› datasets on non peer-reviewed research outputs, such as trade and technical journals
› statistics on the technology balance of payments
› data on activity in high-technology and high-value added sectors, such as investment, employment and external trade.

The figure below shows the key science and innovation data sources for New Zealand. Appendix B (page 57) contains details of the key data sources for science and innovation in New Zealand.

Relevant Data Sources

Funding agencies

Application and contract data

- Administrative data from grants management systems (MBIE, HRC, RSNZ, TEC, MPI, Callaghan Innovation)
- TEC data for PBRF and CoREs

Funding agencies

Reporting and evaluation data

- Annual and final reports from grant recipients
- Evaluation reports
- Stakeholder surveys

Research providers

- Annual reports from universities and research organisations
- Internal information systems
- Tech transfer offices

Official statistics

- R&D Survey
- Business Operations Survey

Bibliometric data sources

- Scopus & SciVal
- Web of Science & InCites
- Altmetric Datasets

International

- OECD
- World Bank
- World IP Organisation (WIPO)

Intellectual property offices

- Patent applications, copyrights, trademarks, other IP

Education sector

- Schooling data
- Tertiary qualification data

In addition to these key data sources, which are dedicated to science and innovation, a number of other data sources are important for answering some of the higher-level questions around the socioeconomic impacts of the innovation system. Statistics New Zealand holds two of the most important databases:

- **The Longitudinal Business Database (LBD)**, which is a linked dataset covering a range of information on New Zealand businesses, including production, employment, trade and tax data. The LBD also contains data from the R&D Survey and the innovation module of the Business Operations Survey.
- **The Integrated Data Infrastructure (IDI)**, which includes individual tax data, education and migration data, births and deaths, and Census data. The IDI is linked to the LBD through the Linked Employer-Employee Data.

Linking science and innovation data, particularly data on funding and researchers, to these national data infrastructures will be critical to answering some of the higher-level enduring questions. The LBD will be important when considering economic outcomes, such as firm and industry-level productivity and firm creation from scientific investment. The IDI will be important when considering the effect of research on the development of human capital and on the labour market. It may potentially also help inform questions on researcher mobility and researcher demographics.

For non-economic outcomes, data holdings of the Ministry of Health, the Ministry for the Environment and the Ministry of Social Development are important. The specific attribution of New Zealand research to social, health and environmental outcomes is challenging and would likely involve a mix of qualitative and quantitative methods.

7 Enduring data needs

Answering the enduring questions requires particular data. This domain plan expresses those data needs in specific questions. These data questions form the heart of the domain plan. Each question is given a ranking 'need to address' derived from two factors:

› the importance of the data to the enduring questions
› how fit-for-purpose the current data is.

The rankings were assigned through a collaborative process involving key stakeholders in New Zealand's research, science and innovation system. Appendices C and D contain details of the scores.

NEED TO ADDRESS	CRITERIA
Very high	Few or no information needs are met. Very strong links to the enduring questions and large and significant data gaps identified.
High	Important information needs are not met. Very strong or strong links to the enduring questions and some significant data gaps identified.
Medium	Some information needs are not met. Some links to the enduring questions and some data gaps identified.
Low	Most information needs are met and those that are not are not of major importance, ie few links to the enduring questions.

In order to address the enduring questions, data needs to be cut in a variety of ways. For instance, for the research system we need to know about funding contracts, funding mechanisms, people, institutions, types of research, research projects, research outputs, and research outcomes.

We have grouped the data questions into eight topics, each of which is discussed below. These topics are guided by the domain plan framework.

Data Need Topics

To answer the enduring questions we need to know about:

1. R&D funding and expenditure
2. R&D outputs
3. People and skills
4. Business R&D and innovation
5. Collaboration
6. Knowledge transfer and commercialisation
7. Infrastructure
8. Costs

7.1 R&D funding and expenditure

Data on funding and expenditure is critical to understanding the R&D system. These data needs cover who funds R&D and who performs R&D. This input data is critical for:

› calculating returns on investment (enduring question 3)
› tracking funding flows through the system and determining the relative benefits of various funding mechanisms (enduring question 4)
› understanding who is performing the research (enduring question 4)
› establishing the best role for government in the system (enduring question 4)
› understanding collaboration across the system through better data on the current research landscape (enduring question 4)
› analysing efficiency and effectiveness of various types of research (enduring questions 3 and 4)
› linking researchers and their activities and outputs to funding sources (enduring questions 3 and 4).

The following questions highlight the specific data needs in this topic along with their priority rating.

DATA QUESTION	NEED TO ADDRESS	LINK TO ENDURING QUESTIONS	CURRENT DATA FIT-FOR-PURPOSE
Data Q1 What research is being done in NZ? What is the breakdown of total R&D expenditure (performing measure) on an annual basis by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) performing institution; (vi) region where research conducted?	Very High	Very High	Low
Data Q2 How much does the Government provide for R&D in total, and how is this split by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) research theme or government goal (including Vision Mātauranga); (vi) funder, funding programme, funding mechanism; (vii) performing institution, research provider type, researcher; (viii) region where research is conducted; (ix) benefitting region?	Very High	Very High	Low
Data Q3 How much does the private sector (business, IROs and private non-profit) spend on R&D in total, and how is this split by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) performing institution and industry sector; (vi) region where research is conducted?	High	High	Med
Data Q4 How much funding for R&D comes from overseas (other governments, international organisations, foreign firms) in total, and how is this split by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) performing institution; (vi) region where research is conducted?	Med	Med	Low
Data Q5 How is government funding for R&D projects actually spent (including against its designated purpose)?	Low	Low	Low
Data Q6 How much co-funding is leveraged for particular contracts, researchers, funding programmes, sectors, disciplines or research to particular socioeconomic objectives?	Med	Med	Low
Data Q7 What support ($) does government provide directly for hubs and precincts as well as incubators?	Low	Low	Med

DATA QUESTION	NEED TO ADDRESS	LINK TO ENDURING QUESTIONS	CURRENT DATA FIT-FOR-PURPOSE
Data Q8 What are the funding sources of R&D performed in CRIs and universities by (i) government sources; (ii) non-government domestic sources; and (iii) overseas?	Med	Med	Low
Data Q9 How much are industry sectors required to spend on R&D into specific topics (due to legislative requirements)? How much levy funding is spent on R&D and in which industry?	Med	Med	Low
Data Q10 What is R&D as a percentage of revenue or added value in each industry (R&D intensity) and for each institution?	Med	Med	High
Data Q11 What are the characteristics of researchers and organisations applying for research funding? For each fund, who are the principal investigators and proposed contracting research organisations and what is the amount of funding sought? Science leader/principal investigator broken down by: (i) field of research; (ii) age; (iii) gender; (iv) ethnicity; (v) affiliated institution and type; (vi) qualification details; (vii) career stage. Research proposal data broken down by: (i) industry sector; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) research theme or government goal (including Vision Mātauranga); (vi) named researchers (proposed FTEs) and their IDs (Scopus, ORCID, NSN); (vii) proposed collaborations; (viii) length of project(s)?	Very High	Very High	Low

7.2 R&D outputs

This topic contains only one question, but the question is broad. R&D outputs are critical to all of the six policy questions, in particular questions 1 and 2 on impact, question 4 on system design and question 6 on the role of people.

The science of science policy relies heavily on R&D output measures to assess the effect of R&D on higher level outcome measures and the effectiveness of various funding mechanisms. Accounting for R&D outputs is also important for transparency purposes. In addition, output data is an essential component of evaluation frameworks and results-based management.

DATA QUESTION	NEED TO ADDRESS	LINK TO ENDURING QUESTIONS	CURRENT DATA FIT-FOR-PURPOSE
Data Q12 What is the quantity and quality of R&D outputs of NZ-based researchers by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) research theme or government goal (including Vision Mātauranga); (vi) funder, funding programme, funding mechanism; (vii) performing institution, research provider type; (viii) researcher characteristic; (ix) region where research is conducted?	Very High	Very High	Low

7.3 People and skills

This topic contains questions on researchers and their mobility, STEM skills, post-study employment outcomes, and managerial and entrepreneurial skills. The data is specifically intended to answer enduring question 6 on the role of people and skills in the research, science and innovation system.

DATA QUESTION	NEED TO ADDRESS	LINK TO ENDURING QUESTIONS	CURRENT DATA FIT-FOR-PURPOSE
Data Q13 What research capacity and capability exists in NZ? How many active researchers are there in NZ and how are they split by: (i) institution, industry sector, research provider type; (ii) career stage; (iii) field of research; (iv) qualification details (the level, place of study, and field of qualification); (v) funder, funding programme, funding mechanism; (vi) research project; (vii) age; (viii) gender; (ix) ethnicity?	Very High	Very High	Low
Data Q14 What STEM skills exist in NZ and what are the trends over time?	Low	High	Very High
Data Q15 What is the involvement and achievement of NZ students in STEM subjects? How is this changing over time? How is this split by: (i) age; (ii) gender; (iii) ethnicity; (iv) decile; (v) rural/urban split?	Low	High	Very High
Data Q16 What are the post-study employment outcomes and career trajectories for Masters and PhD graduates and what does this look like by: (i) age; (ii) gender; (iii) ethnicity?	Low	High	Very High
Data Q17 What is the mobility of NZ's researchers and what are the trends? What proportion and type of researcher (by ANZSRC field of research category) stay in NZ, go overseas, or return? What is the movement within the university and CRI sector and between public institutes and the private sector?	High	High	Low
Data Q18 What is the level of managerial and entrepreneurial skills in NZ?	Med	Med	Low
Data Q19 Who are the top and emerging researchers?	High	High	Low
Data Q20 Who supervises research students?	Low	Low	Low

7.4 Business R&D and innovation

These questions relate particularly to enduring question 5 on government support for business R&D and innovation. Understanding the levels and rates of business R&D is also critical to informing policy question 4 on government investments in public science. An important policy consideration is ensuring public investment in R&D is complementary to private sector R&D and does not crowd out the private sector.

DATA QUESTION	NEED TO ADDRESS	LINK TO ENDURING QUESTIONS	CURRENT DATA FIT-FOR-PURPOSE
Data Q21 What is the absorptive capacity of NZ business? What is BERD on an annual basis and what is the breakdown of BERD by: (i) firm size and by distribution (eg top 25, top 100 R&D spenders); (ii) industry; (iii) firm type; (iv) grant/non-grant recipients; (v) age of firm; (vi) length of R&D project?	High	High	Med
Data Q22 What are the innovation rates of NZ firms by: (i) firm size and by distribution; (ii) industry; (iii) firm type; (iv) grant/non-grant recipients; (v) age of firm? How much do firms spend on innovation?	Low	High	High
Data Q23 How much does the Government spend on providing direct support to business innovation, excluding R&D?	Med	Med	Med
Data Q24 What are the characteristics (eg investment in capital, level of exports, level of collaboration) of firms performing R&D and innovative firms? What is the performance (as measured by revenue, value-add, productivity) of firms performing R&D?	Low	High	High
Data Q25 What are the factors driving firms' decisions on investment in R&D and who should perform it? What are the barriers to increasing investment in R&D?	Med	High	High

7.5 Collaboration

The innovation literature emphasises that innovation rests on strong connections and linkages between the actors in the system. Collaborations and partnerships may be formally developed through agreements and contracts, but many are informal in nature. The connections can take multiple forms:

1. between researchers themselves
2. between research organisations
3. across disciplines
4. between researchers and industry
5. between researchers and other end users.

More and more scientific breakthroughs are resting on interdisciplinary research, and tight connections between researchers and end users are essential to generating value from science and research.

This understanding informs the following two data questions on this topic.

DATA QUESTION	NEED TO ADDRESS	LINK TO ENDURING QUESTIONS	CURRENT DATA FIT-FOR-PURPOSE
Data Q26 What is the extent, nature and quality of collaboration between researchers, either in NZ or offshore (eg collaboration on joint R&D and innovation projects, co-authorship, co-invention in patents, contract R&D, exchanges, conference/event attendance)? Who is doing what with whom?	High	High	Med
Data Q27 What is the extent, nature and quality of collaboration between researchers and end users (cut at project level), and between research institutions and end users?	High	High	Low

7.6 Knowledge exchange and commercialisation

Knowledge exchange underpins the generation of real-world impacts from scientific endeavour and results. While this process of knowledge diffusion is not well understood, the internet and new models of network analysis are enabling the generation of new measures of knowledge exchange and diffusion. Commercialisation is a particular form of knowledge exchange, whereby actors in the system, in particular industry, are willing to pay for the results of science and research.

DATA QUESTION	NEED TO ADDRESS	LINK TO ENDURING QUESTIONS	CURRENT DATA FIT-FOR-PURPOSE
Data Q28 How much resource do universities and research institutes allocate to knowledge exchange activities?	Med	Med	Low
Data Q29 What are the citation rates of research outputs, including publications and patents, and who is doing the citing (cut by individuals, institutions, field of research, industry sector, domestic/international)?	High	High	Med
Data Q30 What are the adoption rates or 'use' rates of research outputs and knowledge exchange activities of researchers and scientists, such as participation on panels, round tables and assessments, engagement with policymakers in central and local government?	High	High	Med
Data Q31 How much direct revenue is generated from the commercialisation of NZ research (eg patenting, licensing, spin-outs) and in what industry and field of research? Which researchers/research providers were responsible?	High	High	Low
Data Q32 What is the quality of commercialisation skills of NZ researchers, research providers and businesses?	Med	Med	Low
Data Q33 What is the level and types of engagement/ communication between the science sector and the public?	Low	Low	Low
Data Q34 What are the levels of engagement between researchers and Māori?	Med	Med	Low

7.7 Infrastructure

Research infrastructure is playing an increasingly important role in the advancement of knowledge and technology. Research infrastructure refers to facilities, resources and related services used by the scientific community to conduct research in their respective fields. Infrastructure may be a single resource at a single location, a network of distributed resources, or virtual. Examples include large-scale research installations, collections, special habitats, libraries, databases, biological archives, clean rooms, data infrastructure, a national research information system, research vessels, coastal observatories, telescopes, synchrotrons and accelerators.

Much of the administrative data on New Zealand research infrastructure exists, but it is kept in separate systems.

DATA QUESTION	NEED TO ADDRESS	LINK TO ENDURING QUESTIONS	CURRENT DATA FIT-FOR-PURPOSE
Data Q35 What is NZ's physical infrastructure for research and what is its financial value?	Med	Med	Low
Data Q36 What is the demand for research infrastructure from different sectors?	Med	Med	Low

7.8 Costs

An efficient system requires costs to be managed and to be as low as possible. A significant portion of the costs incurred are transaction costs, which include time spent in managing institutions, managing funding, applying for funding and reporting. Transaction costs are an important consideration for enduring question 4, in particular in designing funding mechanisms. Efficient research information systems and national data platforms are critical to reducing transaction costs.

Price changes in the research sector can be markedly different from changes in other sectors of the economy.[25] This needs to be taken into account when considering funding allocations. The Australian Bureau of Statistics has developed R&D deflators[26] and the US Department of Commerce updates annually the Biomedical Research and Development Price Index which is used by the National Institutes of Health to maintain purchasing power.[27]

DATA QUESTION	NEED TO ADDRESS	LINK TO ENDURING QUESTIONS	CURRENT DATA FIT-FOR-PURPOSE
Data Q37 What are the costs of doing research (ie inputs) in NZ?	Med	Med	Low
Data Q38 What are the transaction costs in the research system, including governance costs of particular funding mechanisms?	Med	Med	Low

25 For instance, the National Institutes of Health Biomedical Price Index has been higher than the consumer price index for almost every year since the biomedical price index has been running.

26 Australian Bureau of Statistics (2015).

27 US Department of Health and Human Services (2015).

8 Analysis

The previous section listed the data needs, formulated as questions. It also showed the importance of the data, how fit-for-purpose current data is and accordingly the need to address each data question. Those data questions identified as very high and high priority are the focus of this domain plan. This section discusses these key data gaps.

Overall, significant data gaps exist on the inputs and outputs of research, specifically funding and expenditure, research outputs and people and skills. Without this data, evaluation and impact analysis will remain problematic, and transparency of the system will be limited. The emphasis in this domain plan is, therefore, on improving this basic underpinning data.

There are vast collections of data on proxy impact indicators, such as economic production, firm productivity, water quality, climate change, biodiversity or disease burdens. The issue is not collecting more impact measures, but on attribution back to research outputs which can be associated with particular inputs and policy levers. Input and output data provides the backbone for attribution analysis, which may involve quantitative modelling or more qualitative methods such as case studies. Intermediate outcomes of research and science tend to be something like improved knowledge on a particular area. Data to answer whether such an outcome has been achieved is difficult to capture and systematise, but at a minimum the research outputs are needed to inform evaluation of the intermediate outcomes.

8.1 R&D funding and expenditure – Who's doing what research? Who's funding the research?

The biggest data gap identified through the domain plan process relates to the research landscape – data questions 1, 2 and 11 are particularly relevant. Government agencies, the research community and business organisations almost unanimously highlighted this as the most pressing data gap.

At present there is no dataset of research activity and funded research in New Zealand. Unlike many other countries, New Zealand does not have a national research information system platform.[28] Data of funded research cannot currently be aggregated across funders due to a lack of consistency in what information is collected and how it is collected. There is significant variation between the data requirements of funders and the level of data collected. The information requirements across funding mechanisms mean that there is no consistent unit of analysis which would facilitate aggregation.

Government vote appropriations and the R&D Survey provide high-level figures for government investment in R&D. However, the appropriations data only contains an overall figure for a particular fund and the data is constrained by shifts in vote design and changes in funds.[29] This information source is thus of limited use, particularly for time series analysis. The R&D Survey provides slightly more disaggregation of government funding, but it remains high-level, contains lags (up to two years) and is self-reported. More accurate and granular information on government funding could come from real-time administrative data which is linked to various funders' datasets and possibly data of research providers for devolved funds.

28 Refer to page 12 for international examples.

29 For instance, the Public Good Science Fund and the New Economy Research Fund no longer exist. Other funds, such as the Marsden Fund have been more long-standing. Over time, new funds, such as the National Science Challenges and the Primary Growth Partnership, have been established.

Research is becoming more collaborative over time, with funding mechanisms increasingly encouraging multidisciplinary research and connections with end users. In New Zealand, funding for research has become progressively more diverse through new funding mechanisms administered by various agencies. The changes have included more devolved decision making to institutions, such as CRI core funding, and to collaborations of researchers, such as Centres of Research Excellence (CoREs) and National Science Challenges. Another variation is a mechanism, such as that within the Primary Growth Partnership, that involves end users – private or community organisations – deciding how to use research to deliver part of what they are contracted with government to deliver. This devolution of decision-making has resulted in greater information needs in the sector on what and who is being funded. Funding data needs to be brought together to understand the nature and extent of collaboration occurring across the science system. Data needs to enable comparison across the various funds to ensure effective design of the government's investment in the system.

Cost-benefit and impact analysis require identifying a set of inputs and associating them with outputs and outcomes, or beginning with certain impacts and tracing those back to inputs. We therefore need funding data that can be cut in multiple ways so that various groups of research projects can be sensibly compared with different groups of outcomes or impacts. For instance, to examine the effect of research on renewable energy, we need to identify energy research, either through relevant projects or researchers, or a combination thereof. Given that impacts from science are often long term, we need stable longitudinal data that is underpinned by consistent standards. When examining institutional performance, a complete map of funds received through various mechanisms is required. This requires subcontracting to be taken into account.

Effective policy decisions and good investment decisions by funding agencies and devolved entities require knowledge of where the money is currently going and examination of what previous investments have yielded. Tackling complex challenges, such as climate change and food security, requires the exploration and coordination of a variety of diverse research areas. A key requirement for this is the ability to profile research areas (or 'research portfolios') across the system. Identifying research portfolios is a growing trend in funding agencies and research institutes.[30] A consolidated database of research funded, with public access, would lead to more informed decision-making by improving the uniqueness and complementarity of research activity.

A common unit of analysis is needed to structure the data on research activity and funding. At present, data tends to be collected at the contract level. Some contracts are for particular activities or discrete projects, others are for programmes of work, and others are for institutions which then decide how to spend the funding received. Aggregation of this data is therefore challenging due to the very different levels of data collection. Aggregating across funding mechanisms requires a common approach to the profiling of research. The Australia and New Zealand Standard Research Classification (ANZSRC) was developed for this very reason, but its uptake has so far been limited.

In terms of level of analysis, stakeholders indicated that the activity level is too detailed and that collecting information at this level is not practical. Nor are funding agencies in general interested in this level of detail. The project level is the most useful level for both research institutions and funders. Researchers tend to organise themselves into projects and this level is more amenable to description and to meaningful profiles by types of research, etc. Attribution of researcher time spent on a particular project would require estimation given that researchers often work on multiple projects. The next level higher, a programme or a collaboration such as a National Science Challenge, typically contains many different types of research and many researchers, and is thus more problematic to profile and to analyse. Projects, on the other hand, can typically be profiled and can be grouped together into research portfolios and programmes. An exception to this is where research activity forms only part of the contracted project. In this case, sub-project level granularity would be required to determine R&D investment.

30 Wallace and Rafols (2015).

What is a Research Project?

The Frascati Manual (2.12) defines an R&D activity as "the sum of actions deliberately undertaken by R&D performers in order to generate new knowledge". An example of an R&D activity might be carrying out blood tests for patients taking a new drug, or observing animal behaviour in a particular location(s).

When discussing types of R&D, the Frascati Manual tends to use the project as the unit of analysis. According to the manual (2.12), an R&D project:

- **comprises a set of R&D activities**
- **is organised and managed for a specific purpose**
- **has its own objectives and expected outcomes.**

A project generally has:

- **its own budget**
- **a lead researcher**
- **well-defined research questions and methods.**

Because of these characteristics, projects can be profiled by field of research, socioeconomic objective, collaboration with end users, connections with Māori and benefitting region.

An example of an R&D project might be an investigation of a particular mortality to establish the side effects of certain cancer treatments; investigating new methods of measuring temperature; or modelling a crystal's absorption of electromagnetic radiation.

R&D programmes are groups of projects packaged together to form a coherent whole towards a more strategic, long-term vision. Because they tend to be larger and more complex in nature, programmes are more difficult to profile.

Example One: "Develop tools for the management and restoration of aquatic taonga species", $200,000 per year, CRI core funding project, NIWA

Example Two: "Carbon storage in soil and biomass: develop a new method for assessing soil carbon stock changes and develop a new approach to measure net carbon exchange by forest canopies growing in complex terrain", $270,000 per year, CRI core funding project, Landcare

Example Three: "Evolutionary drivers of ornamental colouration in male and female birds", $775,000, 2015 Marsden Grant

Example Four: "Improving care processes for patients with possible acute coronary syndrome", $349,966 over 18 months, Health Innovation Partnership

Example Five: "Model-based diagnostics and therapeutics in pulmonary medicine", flagship project 1 of MedTech CoRE

Example Six: "Undertake research on the natural traits of milk and how they can be maintained in specially processed milk to enable benefits such as reduced risk of developing wider allergies", $1,000,000, High-Value Nutrition National Science Challenge, AgResearch and Miraka

Data capture at the project level should be applied to all funding mechanisms that fund research – even those that are not purchasing projects directly, or where purchased projects only partly involve research activity. The PBRF should be excluded as it is intended to reward research excellence, not to fund research outputs directly. Likewise, it would not be appropriate to collect project-level information for firm R&D grants. Most of these grants are rules-based and the funding provided accounts for only 20 per cent of the R&D the firm performs. Collecting project-level information for devolved funds, such as CRI core funding, CoREs or National Science Challenges, is not designed to question the decisions of these devolved decision makers. It is simply designed to improve transparency, facilitate comparison across funds and to generate time series data across the research system.

Application data is critical to developing a counterfactual for evaluation purposes, ie researchers not funded can be used as a control group against those who received funding, comparing each group's outputs. The demographics of researchers, such as career stage, may affect outputs. Constructing counterfactuals for evaluations therefore needs to take into account these demographic elements. Accordingly, application data needs to include demographics of researchers. A common government CV, used by all funding agencies, could capture this information once.

The question of how government money for R&D is actually spent (data question 5) is important, but of lesser immediacy than the research landscape questions. Regular audits of institutions currently provide the assurance that money is being spent according to its designated purpose. Funding agencies could potentially agree on a set of categories of spending which could inform their investment decisions and policy issues on overheads and infrastructure.

An important measurement issue for R&D statistics is estimating time spent on research in the higher education sector and in the health sector. Measurement of R&D for these two areas is highly dependent on R&D coefficients that are applied to personnel costs. As mentioned in Section 3, it can be difficult to separate teaching and research activities of universities, but separation is required for the measurement of higher education expenditure on R&D. The coefficient was last reviewed in 2010 and the revised Frascati Manual discusses the use of time-surveys and/or administrative data. In teaching hospitals, the activities of teaching, R&D, advanced medical care and routine medical care are frequently closely linked. A consistent approach to measuring R&D in District Health Boards needs to be developed to improve measurement of health-related R&D.

8.2 R&D outputs – What's being produced?

The second key gap identified through the domain plan process related to the results of research (i.e. data question 12). In line with the OECD definition of results, the category of results includes outputs, outcomes and impacts.[31] As indicated previously, data on socioeconomic impacts is plentiful. The attribution back to research and science is more problematic and rests on various methodological techniques. However, all techniques require an understanding of outputs, and rates of return require input data. The first place to start in improving data on the results of research is therefore to capture data on research outputs.

Bibliometric datasets capture a significant portion of research outputs, but not all, particularly in the arts and humanities.[32] They also only capture peer-reviewed publications. Important outputs, such as end user reports or information materials with research findings, are excluded. In New Zealand, there is currently no repository or dataset of research outputs, and funding agencies and research providers do not integrate their reporting data on outputs. Not all funding agencies record the details of each research output and the categories vary, which increases transaction costs for researchers and inhibits system-wide analysis. Capturing all research outputs is critical to evaluation and impact analysis as benefits may otherwise be underestimated.

Researchers receive funding from multiple sources and often have to report through multiple channels. It can therefore be misleading to assign a research output to only one funding mechanism. This can result in less robust evaluation conclusions for those evaluations considering one funding mechanism in isolation. The evaluation may wrongly assume that the output was generated exclusively from the particular fund being evaluated.

---

31 OECD (2002).

32 The reach of these datasets is, however, increasing. For instance, there are emerging datasets on research data, such as Thomson Reuter's Research Citation Index. Datasets are being published and issued digital object identifiers (DOIs).

In recent years, a flurry of small start-ups establish themselves which specialise in research information systems have been established. These systems can quickly profile research based on semantic tools and they enable automatic reporting of research outputs through to funding agencies. Transaction costs of collecting the details of each research output are almost eliminated.[33] Internationally, there has also been a trend towards open access to research outputs, which enhances knowledge diffusion and can improve the quality of research.

8.3 People and skills

The third key data gap that emerged through the domain plan process related to data on research capacity and capabilities in New Zealand. The R&D Survey provides internationally comparable statistics on the number of researchers in New Zealand and several breakdowns of the data. However, more detailed data, such as capacities in particular fields, career stage or researcher demographics, is unavailable. Funding agencies often have detailed information on the researchers they fund, but do not collect the same demographic information on researchers. Furthermore, data from the funding agencies is not linked together. Linking of data could enable a deeper understanding of the capacities of New Zealand's research and science system.

Researchers are at the heart of the scientific process (see diagram on the following page). Data on the research and science system therefore needs to ensure that the researcher is a common thread connecting various datasets. Research outputs, whether collected in bibliometric datasets or by institutional repositories, are tied back to specific researchers. Linking input and output data therefore requires that funding data capture all researchers funded. This will enable the automation of much reporting, as funders can access the records of the researchers funded, provided that the researcher allows the funding agency access. The ORCID system is designed with such functionalities and facilitates the connecting of funding, researchers, and their outputs (see Appendix E page 84 for further information).

Basic Sketch of the Research and Innovation Process

33 Because of concerns over transaction costs, the Foundation for Research, Science and Technology (FRST) and MBIE contestable funding have not previously asked for the details of each publication, just counts. During discussions on the domain plan, several research providers noted that they collected details of research outputs and that funders should ask for these details. This unit record data could be used for analysing the quality of research outputs.

Knowledge diffusion needs to occur in order to generate socioeconomic outcomes. A key mechanism by which this occurs is likely to be movement of researchers.[34] Data question 17 was ranked high because of the likely importance of this transfer mechanism. Research may also increase the quality of teaching, which is critical to the knowledge transfer process. In addition, postgraduate students performing research, under the supervision of leading researchers, are also recipients of the knowledge diffusion process. The existing data infrastructure to analyse these effects on certain populations using various measures, such as earnings or firm productivity, is available through Statistics New Zealand Integrated Data Infrastructure (IDI) and the Linked Employer-Employee Data. However, at present we do not have a way of identifying the set of researchers (including postgraduate students) *and* their funding in the IDI.

Information on the education system, such as the pipeline for STEM skills, the subject choices of students and their career choices and trajectories, is not an area identified as containing significant gaps. Data in this area is comprehensive and is connected into the IDI. This is, for instance, enabling rich data on the relationship between certain qualifications and earnings.

8.4 Business R&D and innovation

The level of BERD is one of the most important indicators of a highly performing innovation system and therefore needs to be monitored on a regular basis with a reliable total figure. As New Zealand's economy moves up the value chain with a greater proportion of high-value and high-knowledge firms and industries, we would expect BERD to increase. As BERD is highly variable across industries due to industry-specific characteristics, industry breakdowns are important to understanding levels, rates and determinants of BERD.

The R&D Survey is the key source of information on BERD along with Callaghan Innovation grant data. As the R&D Survey is a biennial survey an annual figure for BERD through that route is currently unobtainable. Statistics New Zealand produces an estimated BERD figure for those years in which the R&D Survey is not conducted, but the current estimation technique yields very variable results. The R&D expenditure figures derived from the Business Operations Survey cannot be used for generating an official national measure of BERD.[35] Options for improving the estimation of BERD include significantly enhancing the underlying estimation model which could include using Callaghan Innovation grant data. Another option would be to run an annual survey for businesses.

Over time, the quality of government administrative data on business support for R&D has improved. Callaghan Innovation grant data can be linked with the Longitudinal Business Database (LBD) which enables analysis on the impact of BERD and innovation on firm-level performance. The current data infrastructure also allows us to identify how particular characteristics of firms may affect R&D and innovation and vice versa. However, gaps exist around profiling the types of research undertaken and for what purpose.

The innovation module of the Business Operations Survey provides data on the innovation rates of New Zealand firms. The module broadly reflects current international best practice, although international comparisons are at present hindered by the different population frames included across countries. The module could be improved by incorporating questions around expenditure on innovation and the benefits of innovation.

8.5 Other gaps

Other gaps identified through the domain plan process related to collaboration in the innovation system, commercialisation, infrastructure and costs.

34 Labour market research shows that people moving from one firm to another transfer knowledge.

35 R&D expenditure figures from the Business Operations Survey cannot be aggregated to yield a national measure of BERD. National measurements of BERD require specific methods for sampling.

8.5.1 Collaboration

In thinking about collaboration it is useful to separate (i) collaboration between researchers and (ii) collaboration between researchers and end users.

Research is becoming more and more collaborative – both across disciplines, across institutions and across borders. Collaboration between researchers may include joint R&D and innovation projects, co-authorship, co-invention of patents, staff exchanges, conference/event attendance. International evidence suggests that more collaborative research leads to higher quality research. Also, scientific breakthroughs are increasingly being generated through multidisciplinary research endeavours. However, we currently have limited understanding of the types of collaborations that are most important.

Collaboration between researchers is relatively straightforward to capture through funding data (research team makeup, subcontracts, types of research supported) and output data (such as co-authorship). Accounting for sub-contracting is important to understanding the extent of collaboration across different fields and institutions. Research providers have indicated, however, that some aspects of collaboration are not captured, such as the country of the collaborating institution. Use of organisational identifiers, such as Ringgold[36] or GRID[37] would provide this information.

Collaboration between researchers and end users is critical to the innovation process and to knowledge exchange. Collaboration may take the form of providing funding for a research project, undertaking a research project together, co-designing the research agenda and holding joint knowledge exchange activities such as conferences. Contracted research is also important for understanding links between research institutes and end users. End user data needs to be captured in such a way that it can be linked into the LBD to determine the effects of scientists and researchers working with particular firms and industries. This would enable researchers to compare firms/industries identified as working with researchers with those that do not and to generate quantitative findings on the economic impact of research over time.

Agreement on a taxonomy of researcher-researcher collaboration and researcher-end-user collaboration across funding agencies would facilitate system-wide analysis and comparisons across funds. Agreement on the specific data fields for each collaboration is also needed.

8.5.2 Commercialisation

Commercialisation of research is one tangible outcome measure that can be directly linked back to research. Other outcome measures have more factors contributing to them, making them more problematic for demonstrating the impact of research. Universities and CRIs collect the data, but aggregation is difficult due to a lack of consistency in reporting on the measures. University technology transfer offices recorded data in the 2000s, but this collection was discontinued.

Data on commercialisation would be of more value if it was connected back to researchers and funding sources. This would enable, for instance, funding agencies to determine which types of funds and types of research were more or less likely to generate commercialisation outcomes.

8.5.3 Infrastructure and costs

Data gaps on infrastructure and costs were not rated as a high priority for addressing. This was because data in these areas has fewer links to the enduring questions. However, the data gaps are of sufficient importance that, over the longer term, they should be addressed. The key gaps relate to identifying nationally significant infrastructure assets and their use, as well as improving measurement of transaction costs and input costs for research.

36 ORCID is currently using Ringgold identifiers.

37 Global Research Identifier Database (GRID) is an open source catalogue of the world's research organisations, released by Digital Science.

9 Set of actions

Following the identification of knowledge and data needs, and the gap analysis, the domain plan process identified actions to improve the data on research, science and innovation in New Zealand. Based on discussions across government agencies and with key stakeholders, actions considered feasible and necessary to improve data are listed below, along with indicative timing and sequencing.

In line with the domain plan principles (refer framework diagram), the actions are designed to serve several purposes:

1. Improve data sharing and integration across the innovation system
2. Improve access to data and transparency of the research and science system
3. Improve data quality, reusability and the ability to aggregate data
4. Reduce collection burden
5. Develop new methods and improve current methods and processes.

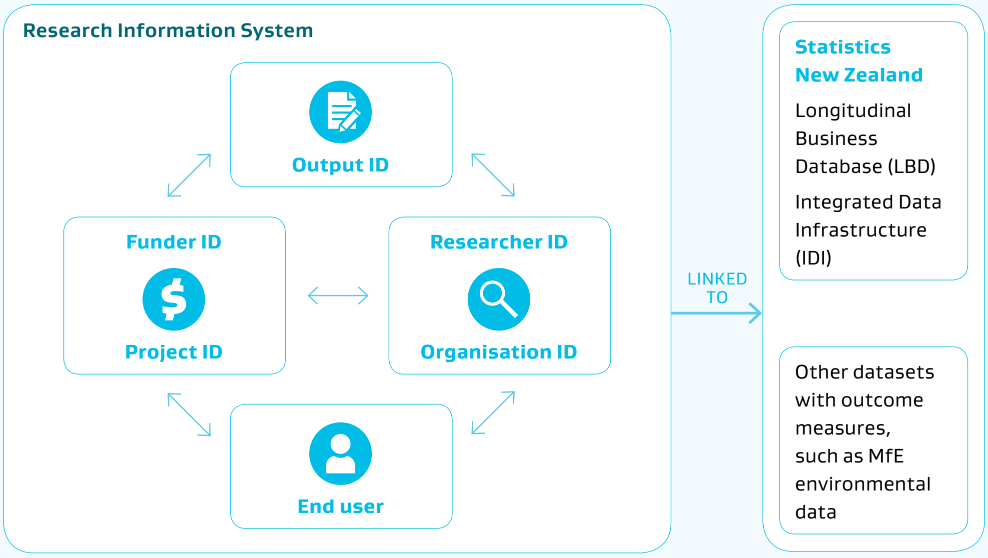
The actions focus on improving data on the inputs and outputs of the research and science system. The focus is not on outcome data or impact measures, as this data already exists in many forms in data holdings elsewhere, such as Statistics New Zealand LBD. The actions will provide a solid base for impact assessment and for improving understanding of the performance of the system. The information will also be critical for investment decisions and for raising the visibility and profile of New Zealand's innovation system.

Implementation of the actions is expected to take at least five years and will require coordinated effort across various government agencies and close engagement with the research community. Key funding agencies and the research community will need to work closely to formulate minimum data standards which will enable the development of linked data on research in New Zealand. Over time, these data standards are expected to feed into application, contract and reporting frameworks and processes. The transition will be relatively straightforward for some agencies and funding mechanisms, but for others implementation may be more challenging and may therefore take several years.

National research information system

ACTION 1	Led by MBIE, funding agencies and the research community establish a **system-wide data information system** for research in New Zealand, ie a national research information system with the ability to link data on researchers, their projects, outputs, funding sources and end user collaboration. This is to be linked to Statistics New Zealand Longitudinal Business Database (LBD) and Integrated Data Infrastructure (IDI) as well as to other datasets with outcome measures.

Building Blocks of National Research Information System



ACTION 2

Led by MBIE, funding agencies and the research community develop a **dataset of research performed in New Zealand and the results of the research** – making as much of the dataset available through a public portal as possible, subject to commercial sensitivities and national security considerations. This would be broadly staged as follows:

i. research funded by MBIE, HRC, RSNZ and Callaghan Innovation, including contestable funds, collaborations (NSCs), science investment funding, R&D grants for businesses, international science funding, awards and fellowships, and support for extension activities such as commercialisation or end-user uptake

ii. research funded by the TEC and R&D and innovation funded by MPI

iii. research commissioned by government departments, Crown entities and district health boards

iv. research performed in Universities and CRIs that is funded by businesses, non-profit organisations, overseas sources and internal funds

v. research funded through industry levies.

ACTION 3

Embed **ORCID identifiers across the research system** by adopting a national approach to ORCID across research providers and funding agencies, which includes (i) funding agencies posting funding information to ORCID profiles or delegating this where practical, and (ii) researchers keeping a complete record of research outputs with acknowledgement of particular funding sources as appropriate, key knowledge exchange activities, peer review contributions and institutional affiliations in New Zealand and offshore.

ACTION 4	Ensure consistent system-wide use of **the Australian and New Zealand Standard Research Classification** (ANZSRC type of activity, fields of research, socioeconomic objective) for researcher and research profiling, and that **searchable project descriptions enable semantic-based profiling of research for ANZSRC, Vision Mātauranga and benefiting region.**
ACTION 5	Statistics New Zealand and MBIE, in conjunction with the Australian Bureau of Statistics and the Australian Research Council, **undertake a review of the ANZSRC** to ensure the classification is fit-for-purpose and reflects changes to research and science since the establishment of the classification in 2008.

Funding agency data

ACTION 6	Funding agencies structure data around **project-level information across all funds**, with the exception of the Performance-Based Research Fund and Callaghan Innovation Growth Grants, and develop a common set of **metadata standards** on the profiling of research (refer Action 4), end user data, domestic and international collaborations, subcontracting, research outputs and knowledge exchange activities to enable aggregation across funds.
ACTION 7	Funding agencies, in conjunction with Statistics New Zealand, collect **end user information** in such a way that facilitates aggregation and connection into other data sources, in particular the Longitudinal Business Database and the Integrated Data Infrastructure
ACTION 8	Funding agencies work on ways to collect accurate information on **research teams for each project in a consistent manner** to facilitate comparison across funds, tracking of research funding and research results, and to allow for automated reporting of outputs and other activities through the use of ORCID records.
ACTION 9	Funding agencies and the research community work on better ways to capture **the use and adoption of research outputs**, making particular use of new methods and techniques available through new and emerging data management technology.
ACTION 10	Funding agencies develop **a taxonomy of budget expenditure categories** across funds to enhance transparency and facilitate comparison.
ACTION 11	Funding agencies develop an **electronic researcher CV** to be used in submitting research funding applications directly to funding agencies, which: i. includes researcher demographic information: gender, ethnicity, age, career stage and qualifications ii. enables automatic data pulling of research outputs and activities from ORCID iii. is fully compliant with New Zealand privacy legislation.

Statistics and surveys

ACTION 12	Statistics New Zealand and MBIE **revise the R&D Survey based on the latest Frascati Manual** to take better account of subcontracting, overseas funding, multinational enterprise R&D, and demographic information on personnel involved in R&D, such as gender and career stage.
ACTION 13	MBIE, Statistics New Zealand and Callaghan Innovation establish a **mechanism to determine a reliable annual figure for business expenditure on R&D (BERD)**, which may involve estimations using Callaghan Innovation grant data, improved modelling of BERD, and/or changes to the current R&D Survey.
ACTION 14	Statistics New Zealand provide **more detailed industry breakdowns of BERD.**

ACTION 15	Statistics New Zealand investigate **discontinuing the biosciences component** of the R&D Survey.
ACTION 16	Statistics New Zealand **broaden the sample of the R&D Survey** to include all district health boards and institutes of technology and polytechnics, and possibly wānanga.
ACTION 17	Statistics New Zealand, in conjunction with MBIE and Universities NZ, **review the R&D coefficient applied in the measurement of Higher Education Expenditure on R&D** (HERD) which will investigate drawing from time-use surveys and administrative data, in line with the latest edition of the Frascati Manual.
ACTION 18	Statistics New Zealand, MBIE and the Ministry of Health, in conjunction with District Health Boards, improve the measurement of expenditure on health-related R&D.
ACTION 19	MBIE and Statistics New Zealand explore **improving the measurement of the benefits of innovation and expenditure on innovation activities.**
ACTION 20	MBIE and Statistics New Zealand **align the firm population sample of the innovation module of the Business Operations Survey** with other comparable surveys internationally to facilitate cross-country comparisons on innovation.

New data collections

ACTION 21	MBIE, in conjunction with the Universities, CRIs, Commercialisation Partner Network members and Callaghan Innovation, develop a **dataset of commercialisation, such as spin-offs and licensing**, which also enables a link to be made between commercialisation activities and the underpinning research and researchers, and funding sources.
ACTION 22	MBIE develop an **inventory of nationally significant research infrastructure** which includes New Zealand's significant databases and collections, and data on use.
ACTION 23	MBIE work with local government agencies to **develop an annual data series of support provided to hubs, precincts and incubators.**

Other

ACTION 24	MBIE and Statistics New Zealand establish a **mechanism for identifying past and present researchers in the Integrated Data Infrastructure and linking in associated funding data** to enable a richer understanding of the effects of research in building human capital and on economic productivity, researcher mobility (domestic and international) and collaboration.
ACTION 25	MBIE develop a means of **determining transaction costs across the research system**, which includes costs in administering funds and time taken to produce an application or reporting document.
ACTION 26	Statistics New Zealand **investigate developing an index of research costs** in New Zealand.
ACTION 27	MBIE and Statistics New Zealand ensure that the **Government Assistance Programme** data in the LBD **includes MPI data and is updated on an annual basis**.

The following figure shows an indicative plan for implementation of the actions. The numbers in brackets relate to the corresponding action.

Indicative Implementation Schedule

FY15/16 | FY16/17 | FY17/18 | FY18/19 | FY19/20 | FY20/21

National research information system & funding agency data

- Define data and collection standards (4,6,7,8)
- Implementation and uptake of standards and project level information
- Establish national dataset for science and innovation (2)
- Embed ORCID into systems (3)
- Capture research output use/ adoption (9)
- Taxonomy for budget expenditure categories (10)
- Develop e-CV and system interfaces (11)
- e-CV rollout

Statistics and surveys

- Revise ANZSRC (5)
- Establish mechanism for annual BERD (13)
- Revise R&D survey (12,14,15,16, 17)
- Rollout new R&D survey 2018
- Revise measurement of innovation expenditure and benefits (19,20)
- Measurement of health related R&D expenditure (18)

New data collections

- Establish dataset for commercialisation (21)
- Data series of support for hubs and precincts (23)
- Inventory of nationally significant infrastructure (22)

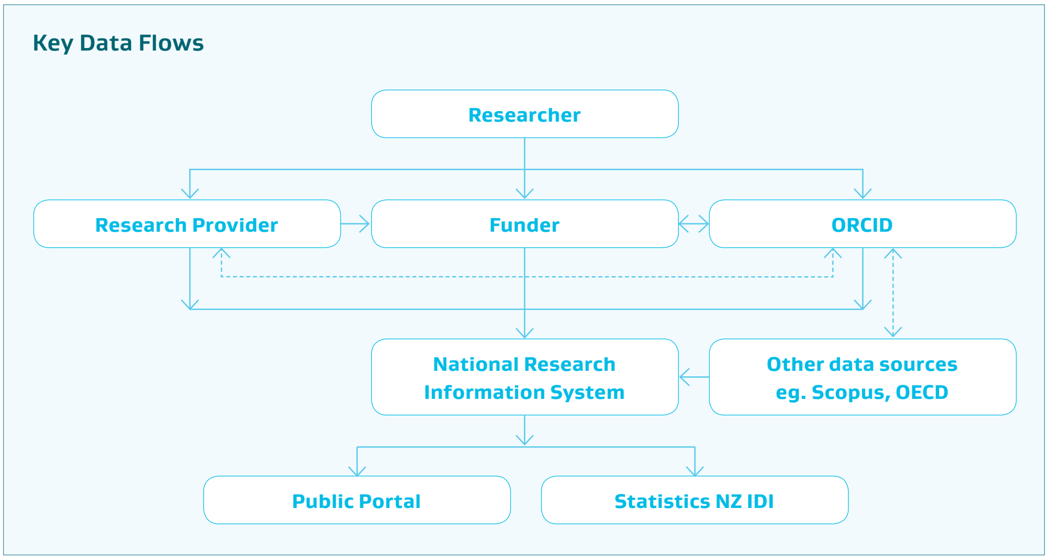
Other

- Identify researchers in IDI (24)
- Update Government Assistance Programme data annually and include MPI data (27)
- Develop means to determine transaction costs across system (25)
- Index for research costs (26)

10 Implications of a national research information system

The most important action emerging from this domain plan is developing a national research information system, a datamart infrastructure, which brings together data on researchers, their projects, outputs, funding sources and end user collaboration.

The figure below shows the high-level key data flows.



The datamart will have various benefits and implications for the different parts of the system. MBIE will be responsible for building the datamart to house the data and for building the public portal. The portal will provide public access to as much of the data as possible, subject to commercial sensitivities and national security considerations.

The public reporting portal will allow users to filter for projects, people, publications or organisations; filter projects by fund, funder, funded amount, field of research, and objective; and filter for projects associated with a given person, organisation or fund. The portal will also have a facility to download data to CSV or excel format and an application programming interface to allow third parties to access the data directly for their own databases and application.

The researcher will:

1. Have information – in one place – on research projects in New Zealand, including associated outputs, and which researchers are working on them
2. Use ORCID records to enable the reuse of data, such as for applying for funding and for reporting purposes
3. Need to ensure their ORCID records are up to date, including details on education, organisational affiliations, activities and outputs
4. Use one CV in applying for funding, which will contain demographic information and connect with the researcher's ORCID record
5. Need to be involved in the process of profiling their research according to the minimum data standards to be developed

6. Likely benefit from a reduction in narrative reporting requirements due to better information on activities and outputs
7. Experience a significant reduction in the variability of information and data requirements for each funding mechanism.

The research organisation will:

1. Have access to past and present research projects from within the organisation and across the sector
2. Be able to search for potential collaborators in New Zealand and pull together more effective research teams
3. Have access to data that enables benchmarking with other organisations, that improves the uniqueness of the institution's research, and that eliminates any duplication in research endeavours
4. Have more visibility of the institution's researchers and research activities domestically and internationally
5. Need to validate ORCID records and outputs for monitoring and reporting purposes, and ensure research outputs contain necessary meta data
6. Need to provide data on research funded by businesses, non-profit organisations, overseas sources and internal funds.

Industry and other end users will:

1. Have the ability to search for research that might be relevant to the firm or organisation
2. Be able to discover researchers on a particular area that might be of benefit to the firm or organisation
3. Have access to the information on research outputs in one place, and if the output is open access, a link to the actual output
4. Have greater visibility of the contribution of industry funding to research undertaken in public research organisations
5. Have the opportunity to supply information on research they fund.

Funding agencies will:

1. Have the ability to compare funding agencies' investments
2. Be able to improve monitoring of research funding
3. Be able to match inputs and outputs unambiguously
4. Need to supply funding data to the central datamart
5. Need to modify systems to incorporate the minimum data standards, which will include information on projects funded
6. Need to post funding information to the ORCID records of awardees.

The system as a whole will:

1. Benefit from greater transparency of research funding, activities and outputs
2. Be in a better position to provide justification to the public of government investment in research, science and innovation
3. Benefit from greater international visibility of New Zealand's research and innovation system
4. Benefit from a reduction in transaction costs in funding, monitoring and reporting processes
5. Be able to generate more meaningful performance measures for the system as a whole
6. Be able to improve comparability of the efficiency and effectiveness of various funding mechanisms.

Glossary of terms

Applied research – original investigation undertaken in order to acquire new knowledge, but which is directed primarily towards a specific practical aim or objective.

Basic research – experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view.

BERD – business expenditure on R&D, ie the total amount of R&D performed by business. This is a measure of R&D performed by business, rather than R&D funded by business.

Bibliometric – the statistical analysis of publications that provides quantitative analysis of academic literature such as comparison of citation rates.

End user – the person or entity that is intended to use, or ultimately uses, the research, science or innovation.

Excellence – well-designed, well-performed, well-reported research, recognised as such, eg through peer review.

Experimental development – systematic work, drawing on knowledge gained from research and practical experience, and producing additional knowledge, which is directed to producing new products or processes or to improving existing products or processes.

GBARD – government budget allocations for R&D, including central, regional and local government.

GERD – gross expenditure on R&D, ie the total amount spent on R&D in a country.

HERD – higher education expenditure on R&D, ie the total amount of R&D performed by the higher education sector, which comprises all universities, colleges of technology and other institutions providing formal tertiary education programmes.

Impact – the direct and indirect 'influence' of research or its effect on an individual, a community, or society as a whole, including benefits to our economic, social, human and natural capital.

Innovation – the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation or external relations.

Mātauranga – knowledge in general, as distinct from mātauranga Māori.

Mātauranga Māori – a body of knowledge that was first brought to New Zealand by Polynesian ancestors of present-day Māori. Mātauranga Māori can exist, and be understood and applied, at various levels, including: broadly by Māori across New Zealand; or at regional, tribal, and whanau levels. Mātauranga Māori can also include the processes for acquiring, managing, applying and transferring that body of knowledge.

Research and development (R&D) – creative and systematic work undertaken in order to increase the stock of knowledge – including knowledge of humankind, culture and society – and to devise new applications of available knowledge.

Research capability – the ability of an organisation, group or individual to conduct a research activity, usually in a specific discipline or field of inquiry.

Research capacity – the amount of research capability possessed by an individual, group or organisation.

Vision Mātauranga – a policy framework whose mission is to unlock the science and innovation potential of Māori knowledge, resources and people to assist New Zealanders to create a better future.

Appendix A: Defining and categorising R&D

The Oslo Manual (2005) defines innovation as "the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation or external relations." The Manual outlines four types of innovation:

1. **Production innovation**: A good or service that is new or significantly improved. This includes significant improvements in technical specifications, components and materials, software in the product, user friendliness or other functional characteristics.
2. **Process innovation**: A new or significantly improved production or delivery method. This includes significant changes in techniques, equipment and/or software.
3. **Marketing innovation**: A new marketing method involving significant changes in product design or packaging, product placement, product promotion or pricing.
4. **Organisational innovation**: A new organisational method in business practices, workplace organisation or external relations.

R&D may or may not be part of the activity of innovation, but it is one among a number of innovation activities. All R&D activities financed or performed by enterprises are considered to be innovation activities. Note that the term 'research' includes the physical and social sciences as well as the arts and humanities. Given the importance of R&D and the international data available on R&D expenditures, it is essential we have a common understanding of what R&D is and is not.

The Frascati Manual has become the standard for R&D measurement worldwide, following initiatives by the OECD, UNESCO, the European Union and various regional organisations. The Frascati Manual (2015)[38] defines R&D as "creative and systematic work undertaken in order to increase the stock of knowledge – including knowledge of humankind, culture and society – and to devise new applications of available knowledge".

For an activity to be an R&D activity, it must satisfy five core criteria:

1. **Novel**: aimed at new findings
2. **Creative**: based on original, not obvious, concepts and hypotheses
3. **Uncertain**: uncertain about the final outcome
4. **Systematic**: planned and budgeted
5. **Transferable and/or reproducible**: lead to results that could be possibly reproduced.

Not all scientific and technological activities are R&D. Non-R&D activities include provision of scientific and technical information, testing and standardisation, feasibility studies, specialised health care and policy-related studies (see Frascati Manual 2.8). Programmatic evaluations may in some cases meet the criteria for an R&D project.

The Frascati Manual outlines three types of R&D:

1. **Basic research**: experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view.

38 See OECD (2015) for the eighth edition of the Frascati Manual.

Basic research can be separated into two further categories:

- **Pure basic research**: carried out for the advancement of knowledge, without seeking long-term economic or social benefits, or making any effort to apply the results to practical problems, or to transfer the results to sectors responsible for their application.
- **Oriented basic research**: carried out with the expectation that it will produce a broad base of knowledge likely to form the basis of the solution to recognised or expected, current or future problems or possibilities.

2. **Applied research**: also original investigation undertaken in order to acquire new knowledge, but which is directed primarily towards a specific practical aim or objective.
3. **Experimental development**: systematic work, drawing on knowledge gained from research and practical experience and producing additional knowledge, which is directed to producing new products or processes or to improving existing products or processes.[39]

Certain activities, while part of the innovation process, rarely involve any R&D. Such activities include patent filing and licensing, market research, manufacturing start-up, tooling up and redesign for the manufacturing process. Prototypes and pilot plants are, however, included in the experimental development category, according to the Frascati Manual. Some design work should be considered as R&D, such as plans and drawings aimed at defining procedures, technical specifications and operational features necessary for the conception, development and manufacturing of new products and processes. Phases 1, 2 and 3 of clinical trials are treated as R&D.

Examples of the types of R&D are provided below:

1. **Natural sciences example**: The study of a given class of polymerisation reactions under various conditions, of the yield of products and of their chemical and physical properties is *basic* research. The attempt to optimise one of these reactions with respect to the production of polymers with given physical or mechanical properties (making it of particular utility) is *applied* research. *Experimental development* then consists of 'scaling up' the process that has been optimised at the laboratory level and investigating and evaluating possible methods of producing the polymer and perhaps articles to be made from it.
2. **Agricultural sciences example**: The investigation of genome changes and mutagenic factors in plants to understand their effects on the phenome is basic research. The investigation of wild potato genomes to locate the genes responsible for resistance to potato blight in an effort to improve the disease resistance in domestic/crop potatoes is *applied* research. Creating a tool for gene editing by using knowledge of how enzymes edit DNA is *experimental development*.
3. **Computer and information sciences example**: The search for alternative methods of computation, such as quantum computation and quantum information theory, is *basic* research. Investigation into the application of information processing in new fields or in new ways (e.g. developing a new programming language or new operating systems) and investigation into the application of information processing to develop tools such as geographical information and expert systems are applied research. Development of new applications for software, substantial improvements to operating systems and application programs are experimental development.
4. **Social sciences example**: Theoretical investigation of the factors determining regional variations in economic growth is basic research. Such investigation performed for the purpose of developing government policy is *applied* research. The development of operational models, based upon statistical evidence, to design economic policy tools to allow a region to catch up is experimental development.

39 The concept of experimental development should not be confused with 'product development', which is the overall process – from the formulation of ideas and concepts to commercialisation – aimed at bringing a new product to the market. Experimental development is just one possible stage in the product development process: that stage when generic knowledge is actually tested for the specific applications needed to bring such a process to a successful end.

5. **Arts and humanities example**: Studying the history and human impact of glacial outburst floods in a country is *basic* research. Using historical records and the techniques of experimental archaeology to recreate an ancient and long-disappeared musical instrument and to determine how it would have been constructed and played is *applied* research. Linguists developing a tool for diagnosing autism in children based on their language acquisition, retention and use of signs is *experimental development*.

Researchers and funding agencies also use the terms '**mission** research' and '**translational** research'. Mission research is similar to applied research in that research is directed towards a specific aim. The implication with mission research is that the specific objective represents a large complex challenge. Translational research aims to make findings from basic science useful for practical applications. It often focuses on multi-disciplinary collaboration and is particularly applied in health research where it is increasingly used as a separate category of research.

The Health Research Council (HRC) has adopted two sets of criteria for classifying research as **translational**. In the first set:

1. The research application is a clinical trial or an intervention, or is observational research aimed at informing policy.
2. The research application demonstrates sustained engagement of stakeholders/end users from the outset (eg patients or communities).
3. The research proposal has the intent of application or uptake, i.e. demonstrated translatability. This needs to be clearly stated and identified within the research proposal.
4. Timeliness (the research is likely to be translated/taken up in the short to medium term).

Alternatively, to be classified as translational, the application must represent a progression from one of the following 'pipeline' categories to another:

1. Discovery of potential prevention/treatment strategy
2. Non-clinical testing, eg in animal models
3. Clinical trial or study to determine efficacy
4. Integration of treatment/prevention strategy into routine practice/real-life situation (effectiveness, cost-effectiveness, etc).

R&D data can be classified according to fields of research. The Frascati Manual proposes the OECD Fields of Research and Development classification which contains six broad categories: natural sciences, engineering and technology, medical and health sciences, agricultural and veterinary sciences, social sciences, and humanities and the arts. Each of these six categories is then further broken down into a total of 42 sub-categories. The Australian and New Zealand Standard Research Classification (ANZSRC) 2008 system[40] contains detailed fields of research (FOR) codes at three levels of disaggregation (two, four and six digits).

R&D expenditures may also be broken down by socio-economic objectives. The ANZSRC contains detailed socio-economic (SEO) categories in defence, economic sectors such as primary production and manufacturing, health, education, law, politics, cultural understanding and environment. Like the ANZSRC FOR codes, the SEO categories contain three levels: two, four and six digits.

40 The Australian Bureau of Statistics and Statistics New Zealand jointly maintain the ANZSRC. See: http://www.abs.gov.au/ausstats/abs@.nsf/mf/1297.0 for further information. The ANZSRC aims to reflect the scope of R&D activity undertaken in Australia and New Zealand; have categories that are mutually exclusive and category names that are as meaningful and precise as possible. It also aims to maximise correspondence with relevant national and international classifications, such as the OECD Frascati Manual and the Australia and New Zealand Standard Industrial Classification (ANZSIC) to facilitate comparisons with other statistics and allow international comparisons.

Appendix B: Details of current data holdings

OFFICIAL STATISTICS	
R&D Survey	The R&D Survey measures the level of R&D activity, employment, and expenditure in the business, government, and higher education (university) sectors. In 2014, the R&D Survey population was 4,500 pre-identified firms performing R&D. Statistics from this survey are used in the development of science policy in areas such as the setting of research priorities, government research funding levels, science education and innovation encouragement schemes. Biennial survey. Data from 1996. http://www.stats.govt.nz/browse_for_stats/businesses/research_and_development.aspx
Business Operations Survey (BOS)	The Innovation Module of the Business Operations Survey (BOS) provides information on the characteristics of innovation in New Zealand firms: i. Levels of firm innovation, including R&D expenditure and development of product and organisational innovations ii. How and why firms collaborate with other firms and institutions in order to innovate iii. Source of ideas or information for innovation iv. Factors affecting the ability of firms to innovate v. Outcomes of innovation for firms, including its effect on exports The innovation module is included in the BOS every two years. Data from 2003 (incorporated into the BOS from 2005). Module A of the BOS covers general questions on business operations and is conducted annually. Module A asks respondents for the amount of expenditure on R&D. http://www.stats.govt.nz/browse_for_stats/businesses/business_growth_and_innovation/innovation-nz-landing-page.aspx
FUNDING AGENCIES	
Funding agencies application, contract and reporting data	This is data held by funding agencies (MBIE, HRC, RSNZ, TEC, MPI, Callaghan Innovation) on applications, awarded contracts and subsequent reporting by recipients. The data provides description of research, researchers involved, contracts and funding, outputs and outcomes. The different agencies collect different data and use different classifications and data standards, thus limiting the ability to aggregate data in any meaningful manner. Includes: i. Grant contract data and annual reports ii. CRI core funding reporting data iii. HRC workforce census iv. MBIE Pre-seed Accelerator Fund reporting data v. TEC CoREs data

Information from evaluation reports	Information collected as part of mid-term, final or post-contract evaluations: i. Post-contract evaluation information (currently applicable for the Primary Growth Partnership and the Royal Society of New Zealand) ii. Process evaluation reports iii. Impact evaluation reports.
Tertiary education research financing statistics	Sources and allocations of research finance in the tertiary education sector. • Vote TEC funding for research by fund type • PBRF allocations by component, subsector • University external research income by source • University PBRF external research income • University research contract income • Expenditure on R&D by universities (2002-2013) • University research expenditure by research type (2005-2013) • University research expenditure by purpose (2009-2013) https://www.educationcounts.govt.nz/statistics/tertiary/research From 2004 onwards - coverage differs depending on statistic
PBRF Quality Evaluation	TEC produces an electronic report based on the results of the Quality Evaluation assessment which includes: • Quality rating of staff at an aggregate level • FTE information at an aggregate level • Nominated and other research outputs by type • Demographic information such as age and gender. 55 per cent of PBRF funding is linked to the Quality Evaluation.
PBRF staff data file	Supports the submission of Evidence Portfolios. Ministry of Education tertiary education organisation staff data collection provides some demographic data, while the TEC collects information to support the staff eligibility audit.
PBRF annual reports	Report on how Tertiary Education Organisations performed against PBRF measures. Includes: • funding received by each TEO • external research income received by each tertiary education organisation • changes in overall funding and funding by source • changes in indicative versus final funding http://www.tec.govt.nz/Funding/Fund-finder/Performance-Based-Research-Fund-PBRF-/Publications/
MBIE surveys	a. Science in Society Survey Measures attitudes of the general public towards science and technology, whether the level of information respondents see and hear about science is too much or too little, level of interest in science and technology, etc. b. CRI Stakeholder Survey/National Science Challenge Survey These surveys provide data for the evaluation of performance indicators used in CRI monitoring and in the NSC performance framework. For example, CRIs have indicators regarding research priorities relevant to sector needs, ability to form appropriate teams to deliver CRIs' outcomes, end-user adoption of knowledge and/or technology which they need to include in their Statements of Corporate Intent. c. Pre-seed Accelerator Fund Outcomes 10-year Review This was a one-off review, conducted in 2014. It involved data collected from 15 New Zealand research organisations that invest the majority of PSAF. https://www.kiwinet.org.nz/files/Investment/Historic-PreSeed-Report.pdf

<table>
<tr><td>University commercialisation offices data</td><td>Between 2003 and 2008, university commercialisation offices produced data on revenue (such as license income, contract research income), new firms, and research productivity (such as invention disclosures, new patents filed).</td></tr>
<tr><td colspan="2">EDUCATION SECTOR</td></tr>
<tr><td>Education Counts data</td><td>MoE publishes a range of data including:
• Secondary school subject enrolments
• School to tertiary transitions – covers Trends in International Mathematics and Science Study (TIMSS)
• Tertiary enrolments and completions
• Post-study employment outcomes
• Staffing data in tertiary education institutions
• Bibliometric data (starting from 1981)
• Commercialisation statistics from UCONZ (2003-2008)
www.educationcounts.govt.nz</td></tr>
<tr><td colspan="2">RESEARCH PROVIDERS</td></tr>
<tr><td colspan="2">Crown Research Institutes</td></tr>
<tr><td>Published annual reports from CRIs</td><td>Reports both financial and non-financial information and includes:
• Information to enable an informed assessment of the CRI’s and subsidiaries’ performance during the financial year, particularly by Ministers, MPs and Select Committees.
• A comparison of the performance of the CRI and subsidiaries against its Statement of Corporate Intent impacts, outcomes and objectives.
NB Performance indicator measures and reporting vary considerably across the different CRIs.</td></tr>
<tr><td>Annual CRI reporting to MBIE</td><td>CRI core funding reporting data to MBIE includes a description of projects, funding, and alignment to industry sector, collaboration and subcontractor aspects of the project.
• Does not include type of research, ANZSRC FOR, researchers, publications, FTE/headcount information by project
• Information provided to MBIE by way of a core reporting spreadsheet. Responses are free format and not necessarily consistent across different CRIs so some work required to standardise if this is to be looked at as an aggregate
• Information collected differs from year to year.</td></tr>
<tr><td>CRI internal information systems</td><td>Several CRIs have developed their own internal budgeting and reporting systems which generate data on expenditure, activities and outputs.</td></tr>
<tr><td colspan="2">Universities</td></tr>
<tr><td>University annual reports</td><td>Contain financial statements, including research income, and data on staff, students and qualification completions</td></tr>
<tr><td>University internal information systems</td><td>University information systems generate data on research outputs and financial information</td></tr>
<tr><td colspan="2">Other</td></tr>
<tr><td>Return on Science data</td><td>Return On Science is a national research commercialisation programme to deliver new research to market from universities, research institutions and firms. These services include connecting science and technology with strategic management, advice and guidance, and access to capital. Dataset includes project name, project stage, principal investigator, client, project type (physical sciences, biotech and life sciences, ICT, agritech and agbio).
http://www.returnonscience.co.nz/</td></tr>
</table>

KiwiNet	The Kiwi Innovation Network (KiwiNet) is New Zealand's network of public research organisations, working together to transform scientific discoveries into marketable products and services. Primarily, Kiwinet comprises universities, CRIs and Callaghan Innovation. The following are available from their website: commercialisation stage, project description, researcher, organisation, patents and plant variety rights. https://www.kiwinet.org.nz
Private research organisation annual reports	Annual reports of private research organisations contain financial accounts and other information on research activities and achievements.
Survey of New Zealand Scientists and Technologists	Surveys carried out in 1994, 1996, 2000 and 2008 (1994 and 2008 by the New Zealand Association of Scientists and 1996 and 2000 by the Royal Society of New Zealand). The surveys contain information on the attributes and accomplishments of the New Zealand research, science and technology workforce, the concerns of scientists, their values relating to science and society, and their opinions on the performance of the science and innovation system. The results of the 2008 survey are available here: http://www.scientists.org.nz/files/journal/2010-67/NZSR_67_1.pdf.
INTELLECTUAL PROPERTY	
Intellectual Property Office	The Intellectual Property Office of New Zealand (IPONZ) is the government agency responsible for the granting and registration of intellectual property rights. Specifically: • patents (applications, patents granted and patent citations) • trade marks • design and plant variety rights. Statistics New Zealand's Longitudinal Business Directory includes patent applications, designs and trademark registrations filed with IPONZ. www.iponz.govt.nz
BIBLIOMETRIC AND ALTMETRICS SOURCES	
Scopus (Elsevier)	One of the major academic literature collections, Scopus is a bibliographic database containing abstracts and citations for academic journal articles. It covers approximately 22,000 titles from over 5,000 publishers, of which 20,000 are peer-reviewed journals in the scientific, technical, medical, and social sciences (including arts and humanities). It is owned by Elsevier and is available online by subscription. Searches in Scopus also incorporate searches of patent databases.
Web of Science (Thomson Reuters)	Like Scopus, Web of Science is also a major academic literature collection. It covers more than 12,000 journals in all subject areas. The collection contains data back to 1900 and is available for 22 broad subject areas and over 250 narrow subject areas.
Altmetric	Altmetric tracks what people are saying about papers online on behalf of publishers, authors, libraries and institutions. Provides article-level metrics based on scanning social media sites, newspapers, government policy documents and other sources for mentions of scholarly articles (eg tweets, blogs or posting of public links). It tracks articles from hundreds of different publishers, preprint databases and institutional repositories. http://www.altmetric.com/ Note : Altmetrics as a *field* is the creation and study of new metrics based on the Social Web for analysing, and informing scholarship. For more on this see www.almetrics.org.

Other data sources, while not specifically about science and innovation, are important to answering the enduring questions.

› The Longitudinal Business Database (LBD) is a linked longitudinal dataset covering a range of business information, including production, employment, trade and tax data. The database covers administrative sources of financial performance, employment and wage data, coupled with a variety of sample surveys and other (notably Customs merchandise trade) data. The LBD includes business demography statistics from the Business Register Update Survey, business tax data, the R&D Survey, the Business Operations Survey, the Annual Enterprise Survey and Overseas Merchandise Trade. It also contains the *Government Assistance Programme* dataset: lists of firms that have received assistance from government agencies, together with information on the nature of the assistance.[41]
› The Integrated Data Infrastructure includes individual tax data, education and migration data, births and deaths, and 2013 Census data.
› The Ministry for the Environment and Statistics New Zealand report on the state of the environment, the pressures that have created that state, and the way the state of the environment impacts other spheres of the environment and our life (public health, ecological integrity, the economy, te ao Māori, and culture and recreation). The Environmental Reporting Act requires publishing of one domain report every six months and a synthesis report on New Zealand's environment as a whole every three years. The five domains are air, atmosphere and climate, land, freshwater and marine.
› The Ministry of Social Development provides information on inequality and hardship, children and young people.
› The New Zealand General Society Survey contains information on the well-being of New Zealanders, social networks, social and civic participation, norms and trust, community support and cultural participation.
› The Ministry of Health collects and publishes data on mortality, morbidity and health risk and protective factors.
› International Trade in Services and Royalties Survey collects data on R&D services.
› Industry organisation levy reports contain information on R&D supported through the levy.
› Annual reports of not-for-profit organisations sometimes contain information on research funded.

41 Agencies supplying data are MBIE (and its predecessor organisations, including the Ministry of Tourism), New Zealand Trade and Enterprise, Te Puni Kokiri and the Ministry of Social Development. The data includes the period in which the firm received assistance, the specific programme accessed and, where appropriate, an estimate of the value of that aid.

Statistics New Zealand
TATAURANGA AOTEAROA

Integrated Data:

Integrated Data Infrastructure content April 2016

Education
- Early Childhood Education Providers
- Primary Education – From 2007
- Secondary Education – From 2004
- Tertiary Education – From 1994

Benefits and services
- Benefit Dynamics Data – From 1990
- Child benefit, and 2nd and 3rd tier benefit payments – From 1993
- Youth Services – From 2006

Personal tax
- EMS – From 1999
- IR3 – From 1997
- IR20 and IR4S – From 1995

Student Loans and Allowances
- From 1992

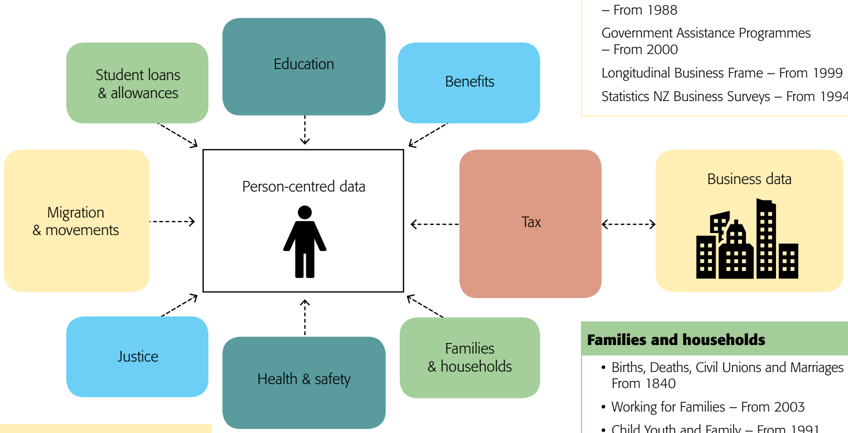
Business Centred Business tax
- IR10 – From 1999
- IR4 – From 1999
- GST – From 1992

Customs – Overseas merchandise trade – From 1988

Government Assistance Programmes – From 2000

Longitudinal Business Frame – From 1999

Statistics NZ Business Surveys – From 1994



Families and households
- Births, Deaths, Civil Unions and Marriages – From 1840
- Working for Families – From 2003
- Child Youth and Family – From 1991
- Tenancy – From 2000
- Household Economic Survey – From 2006
- Household Labour Force and NZ Income Surveys – From 2006
- Survey of Family Income and Employment – 2002-2010
- Social Housing Data – From 1980

Migration and movement
- Border movements – From 1997
- Visa applications – From 1997
- Departure and Arrival cards – From 1997
- Longitudinal Immigration Survey of NZ – 2005-2009
- Migrant Survey – From 2012

Health and safety
- Population Cohort Demographics – From 2004
- NHI and PHO Addresses – From 2004 and 2005, respectively
- PHO Enrolments – From 2003
- Chronic Conditions – Complete from 2007
- Cancer Registrations – From 1995
- Mortality – From 1988
- Pharmaceuticals – From 2005
- Publicly Funded Hospitals Discharges – From 1988
- NNPAC – National Non-admitted Patient Collection – From 2007
- Lab and GMS Claims – From 2003 and 2002, respectively
- NIR – From 2006
- B4School checks – From 2011
- PRIMHD – From 2008

Census
- 2013 Census

ACC
- Claims and medical codes 1994-2015

Justice
- Recorded crime: victims – From 2014
- Recorded crime: offenders – From 2009
- Court charges – From 1992
- Sentencing and remand – From 1998

Statistics NZ operates a 'five safes' framework ensuring safe people, safe projects, settings, data and safe output. Read more about how privacy is managed on our website.

www.stats.govt.nz/idi

International organisations provide a range of science and technology indicators.

i. OECD

a. Science, Technology and Industry Scoreboard
b. Main science and technology indicators (MSTI)
c. STI Outlook
d. ANBERD database
e. Patent statistics, including triadic database
f. Database on Immigrants in OECD Countries (DIOC)
g. Programme for International Assessment of Adult Competencies (PIAAC).

ii. World Bank

World Development Indicators on science and technology

a. IP payments and receipts (IMF Balance of Payments Statistics Yearbook)
b. High-tech exports (Comtrade database)
c. Patent and trademark applications (WIPO)
d. R&D expenditure and personnel (UNESCO)
e. Scientific and technical journal articles (US National Science Board).

iii. WIPO

a. IP databases – patents, trademarks, industrial designs, utility models, Patent Cooperation Treaty (PCT) applications.

Appendix C: Data fit-for-purpose assessment

DATA QUESTION	NEED TO ADDRESS	FITNESS-FOR-PURPOSE	COMPLETENESS	CONNECTEDNESS	CONSISTENCY	RELEVANT COMMENTS FOR FITNESS-FOR-PURPOSE ASSESSMENT
R&D FUNDING AND EXPENDITURE						
Data Q1 What research is being done in NZ? What is the breakdown of total R&D expenditure (performing measure) on an annual basis by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) performing institution; (vi) region where research conducted?	Very High	Low	M	L	L	Data on performing institution is better than by funding source due to the way the Frascati Manual recommends R&D Surveys be organised, i.e. the Frascati Manual statistics are based on the performing entity, not the funding entity. The R&D Survey is the primary source of information to answer this question at a high level. The survey runs biennially, so annual figures not available. The survey provides breakdowns by socioeconomic objective (into 14 high-level categories), research type (basic, applied, experimental development) and industry of performing institution (13 high-level ANZSIC categories), but not field of research, horizons or region where research is conducted. The location of private sector research can be found through linking to the LBD which contains the location of the head office. Further breakdowns of ANZSIC needed to calculate R&D intensities for sub-industries. Administrative data of funding agencies could provide more detail on research performed in New Zealand, but funding agency data is not linked. Use of funding agency data requires all sub-contracting to be taken into account. Sub-contracting is not consistently captured by funding agencies. Information systems of research providers could answer this question, but the lack of standard classifications across providers inhibits the ability to aggregate data. Annual reports of universities and CRIs do not provide breakdowns of their research in consistent ways.

Data Q2 How much does the Government provide for R&D in total, and how is this split by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) research theme or government goal (including Vision Mātauranga); (vi) funder, funding programme, funding mechanism; (vii) performing institution, research provider type, researcher; (viii) region where research is conducted; (ix) benefitting region?	Very High	Low	M	L	L	The R&D Survey provides a total figure for government funding of R&D, but does not provide breakdowns. Funding agencies collect data that allow them to answer most of this question individually. However, a lack of consistent classifications, taxonomies and business rules across the different funding agencies significantly hinders the ability to consolidate data. For example, ANZSRC FOR and SEO are used by some agencies, but not others – some code to 4-digit, others to 6-digit level, some allow for weights to be applied, others do not. Most funding agencies have adopted their own profiling classification which is useful for their own purposes, but inhibits system-wide aggregation of data. Frascati research-type profiling is not done for all research funded, nor is all research profiled for Vision Mātauranga. All funding agencies collect the named investigators, but there is inconsistency on data collected on the whole research team. Information on benefitting region is mostly not collected. Region where the research is conducted could be derived from the affiliation addresses of researchers. The research performing institution is not classified consistently across funding agencies, and no funding agency classifies this by ANZSIC. Organisation IDs are not used. Difficulty in dealing with the life of a contract in that the research may move across classifications, eg from applied research to experimental development. Beyond funding agencies, other government agencies that fund research do not adopt any classifications or profiling that would allow this question to be answered.

Data Q3 How much does the private sector (business, IROs and private non-profit) spend on R&D in total, and how is this split by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) performing institution and industry sector; (vi) region where research is conducted?	High	Med	M	M	M	The R&D Survey is the primary source of data for this question. However, the official statistics have limited ability to answer this question because they are based on performing institution rather than funding source. The R&D Survey captures industry sector for performing firm, not funding firm. Callaghan Innovation grant data is a useful source for inferring firms' funding of R&D. The grant data currently cannot answer most sub-sets of this question (but it is able to answer subset (v) and (vi)). An issue is that the Callaghan Innovation definition of R&D is different from the Statistics New Zealand definition. Industry organisation levy reports and non-profit organisation reports often provide an aggregated figure on research funded, but research is not classified which creates problems for large bodies which fund multiple types of research for various objectives. For private-sector funded research performed in CRIs and universities, could potentially obtain data from CRIs and universities, but currently these institutions do not adopt standard classifications of research projects.
Data Q4 How much funding for R&D comes from overseas (other governments, international organisations, foreign firms) in total, and how is this split by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) performing institution; (vi) region where research is conducted?	Med	Low	L	L	L	The information currently collected through the R&D Survey and the International Trade in Services and Royalties Survey provides an aggregate picture of overseas income. Breakdowns are unavailable. PBRF reporting on external research income is broken down to totals by source, which includes overseas research income. For overseas funded research performed in CRIs and universities, we could potentially obtain data from CRIs and universities, but currently these institutions do not adopt standard classifications of research projects.
Data Q5 How is government funding for R&D projects actually spent (including against its designated purpose)?	Low	Low	M	L	L	Information is not currently collected and collecting it could be problematic. Collection could impose significant administrative burden on some organisations if their accounting systems don't currently track expenditure back to specific funding. There is no agreed classification on budget breakdowns across funding agencies for use in applications. A lack of transparency in the use of overheads has been identified as an issue.
Data Q6 How much co-funding is leveraged for particular contracts, researchers, funding programmes, sectors, disciplines or research to particular socioeconomic objectives?	Med	Low	H	L	L	There is currently no standardised approach for collecting co-funding information. Some funders collect some information, others do not. There is no agreed definition of co-funding across the system (eg direct, related, cash, in-kind, etc) so it is difficult to aggregate information from different funders.

Data Q7 What support ($) does government provide directly for hubs and precincts as well as incubators?	Low	Med	M	M	L	Precincts and incubators receive support from local government and private sector non-profit organisations. No agency consolidates the data. Central government support for incubators is known.
Data Q8 What are the funding sources of R&D performed in CRIs and universities by (i) government sources; (ii) non-government domestic sources; and (iii) overseas?	Med	Low	M	L	L	Note this question is covered in data questions 1-4 above. Lack of consistency across universities and CRIs as to what is included in government sources. Particular issues are dealing with sources from CRIs and dealing with sub-contracting. From 2015, tertiary education organisations participating in PBRF are required to report external research income by New Zealand government contestable funds, New Zealand public sector contract research, overseas research income and New Zealand non-government sources such as industry, iwi and not-for-profit sector.
Data Q9 How much are industry sectors required to spend on R&D into specific topics (due to legislative requirements)? How much levy funding is spent on R&D and in which industry?	Med	Low	M	L	L	Do not currently have a list of legislative requirements for research. Levy funding data is not easily accessible. Industry organisations' reports contain total amount spent on R&D only.
Data Q10 What is R&D as a percentage of revenue or added value in each industry (R&D intensity) and for each institution?	Med	High	M	H	H	Industry revenue and value-add is available in the LBD. R&D Survey reports on industry through 13 high-level ANZSIC categories – further breakdown required if R&D intensities for sub-industries are to be calculated.
Data Q11 What are the characteristics of researchers and organisations applying for research funding? For each fund, who are the principal investigators and proposed contracting research organisations and what is the amount of funding sought? Science leader/principal investigator broken down by: (i) field of research; (ii) age; (iii) gender; (iv) ethnicity; (v) affiliated institution and type; (vi) qualification details; (vii) career stage. Research proposal data broken down by: (i) industry sector; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) research theme or government goal (including Vision Mātauranga); (vi) named researchers (proposed FTEs) and their IDs (Scopus, ORCID, NSN); (vii) proposed collaborations; (viii) length of project(s)?	Very High	Low	L	L	L	Refer also Data Question 13. Information collected about researchers and the research team differs across funders. Not all funders collect details on the research team. Most funders collect gender and ethnicity information. Career stage classifications differ across funders. There is no consistency in collection and classification of field of research for researchers and their proposed research. Application data is not classified uniformly (see issues identified in earlier questions). Not all funders store information on applications in a structured database.

R&D OUTPUTS

Data Q12 What is the quantity and quality of R&D outputs of New Zealand-based researchers by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) research theme or government goal (including Vision Mātauranga); (vi) funder, funding programme, funding mechanism; (vii) performing institution, research provider type; (viii) researcher characteristic; (ix) region where research is conducted?	Very High	Low	M	L	L	Bibliometric datasets are improving in coverage, but they are still not complete datasets of all research outputs. They cover peer-reviewed publications well in the hard sciences and medicine, but less so in the arts and humanities. Various classifications exist for field of research, such as All Science Journals Classification. Linking research outputs to funding is problematic when researchers often receive funding from multiple sources, and outputs can be a culmination of activity from several projects. Some funds have detailed attribution rules for output attribution. A lack of researcher IDs currently inhibits ability to link researchers to their outputs unambiguously. Institutions' datasets of research outputs do not always capture all research outputs of their researchers. Funding agencies have different classifications for outputs. New Zealand funding agencies currently do not require researchers to acknowledge funding sources in their outputs, yet this is common practice overseas. Citations are a proxy of quality, with journal half-life being approximately seven years and books even longer. Significant differences exist between research disciplines in citations, but field-weighted citation measures have improved comparability across disciplines.

PEOPLE AND SKILLS						
Data Q13 What research capacity and capability exists in New Zealand? How many active researchers are there in New Zealand and how are they split by: (i) institution, industry sector, research provider type; (ii) career stage; (iii) field of research; (iv) qualification details (the level, place of study, and field of qualification); (v) funder, funding programme, funding mechanism; (vi) research project; (vii) age; (viii) gender; (ix) ethnicity?	Very High	Low	L	L	L	Refer also Data Question 11. The R&D Survey provides an overall figure of the number of researchers and what sector they are in (government, higher education or business). Contains no demographic information on researchers. There is no register or dataset of active researchers, and individual institutions have difficulty in maintaining up-to-date records. ORCID records could potentially begin to answer some of these questions, but they do not contain demographic information. Census data could potentially provide information by ANZSCO classifications. Currently there is no way of identifying researchers in the IDI, which has qualification data from 1994 and demographic data. The government CV for scientists, developed in 2011, is used by MBIE, HRC and RSNZ. At present the only way to determine researchers associated with different projects is to use funding application and contract data. However, not all researchers funded are captured and there are inconsistencies in demographic information collected. A particular issue is updating research team information when a contract spans multiple years.
Data Q14 What STEM skills exist in New Zealand and what are the trends over time?	Low	Very High	H	M	H	Datasets such as Census and surveys of the adult population are best used for this question. Ministry of Education has unit record data (tertiary education enrolment and completion data) from 1994, but information on the specialisation of graduates that would identify STEM students is only available from 2005.
Data Q15 What is the involvement and achievement of New Zealand students in STEM subjects? How is this changing over time? How is this split by: (i) age; (ii) gender; (iii) ethnicity; (iv) decile; (v) rural/urban split?	Low	Very High	H	H	H	Information on achievement of students in STEM subjects is good and data can be broken down by the categories identified in the question. Population data is available from the July roll return for primary and secondary schools. Each school provides information on the number of students studying each subject. Using the standards that students attempt assessment for on the New Zealand Qualifications Framework (NZQF) trends can be monitored in any population. The NZQF data is available from 2002, but the data defining the population at the student level is suitable from 2008.

Data Q16 What are the post-study employment outcomes and career trajectories for Masters and PhD graduates and what does this look like by: (i) age; (ii) gender; (iii) ethnicity?	Low	Very High	H	H	H	The IDI has reliable matched information for students whose last year of study was 2003 or after. To obtain post-study employment outcomes and occupations we need to link this with Census data, which was incorporated into the IDI in September 2015.
Data Q17 What is the mobility of New Zealand's researchers and what are the trends? What proportion and type of researcher (ANZSRC FOR) stay in New Zealand, go overseas, or return? What is the movement within the university and CRI sector and between public institutes and the private sector?	High	Low	L	H	L	Statistics New Zealand's IDI could be used to determine mobility of New Zealand's researchers but we would need a way to identify researchers. One option is to determine this from occupation information in the Census: ANZSCO classification scientists (234), university lecturers (2421) and health professionals (25). As ORCID records contain institutional affiliation, they could also provide insight into researcher mobility. OECD Database on Immigrants in OECD and non-OECD Countries (DIOC) provides a dataset on the country in which a PhD holder is residing and the country of origin of that individual.
Data Q18 What is the level of managerial and entrepreneurial skills in New Zealand?	Med	Low	L	L	1	Need to better define 'managerial and entrepreneurial skills'. The Innovation Module of the Business Operations Survey (BOS) contains questions on the barriers to innovation, which includes management capability. Business start-ups can be obtained from the LBD. This provides an indication of entrepreneurialism.
Data Q19 Who are the top and emerging researchers?	High	Low	M	L	L	Definitional issues around what is a 'top and emerging researcher', eg amount of funding awarded, publications produced, engagement with end users. Problem with citation lags of publication and identifying researchers early enough. Individual institutions may be able to identify their top and emerging researchers, but at a national level this is more challenging.
Data Q20 Who supervises research students?	Low	Low	H	L	L	No good source of information for who is supervising research students, but research information systems are improving on capturing this information. There are moves internationally to capture this information systematically in researchers' records

BUSINESS R&D AND INNOVATION

Data Q21 What is the absorptive capacity of New Zealand business? What is BERD on an annual basis and what is the breakdown of BERD by: (i) firm size and distribution (eg top 25, top 100 R&D spenders); (ii) industry; (iii) firm type; (iv) grant/non-grant recipients; (v) age of firm; (vi) length of R&D project?	High	Med	Mto H	M	M	We can answer this question quite well from current data sources on the assumption that the level of BERD is a good proxy for the absorptive capacity of New Zealand business. The R&D Survey, BOS and other data in the LBD can answer all parts of this question, the exception being the length of R&D project which is useful to know when designing government grant programmes. BOS industry and sub-industry BERD figures are not suitable for generating industry or sub-industry BERD totals. Need to ensure that the Government Assistance Programme (GAP) data feed, which contains Callaghan Innovation grant data, is complete and loaded into the LBD every year.
Data Q22 What are the innovation rates of New Zealand firms by: (i) firm size and by distribution; (ii) industry; (iii) firm type; (iv) grant/non-grant recipients; (v) age of firm? How much do firms spend on innovation?	Low	High	H	H	H	This question can mostly be answered from the Innovation Module of the BOS and the GAP data for grants. However, we do not currently know how much firms spend on innovation, thus have no intensity measure of innovation.
Data Q23 How much does the government spend on providing direct support to business innovation, excluding R&D?	Med	Med	M	M	M	It can be difficult to separate R&D from non-R&D activities, such as commercialisation, knowledge transfer or extension activities. This is particularly the case for some MPI funds, which fund the whole innovation pipeline.
Data Q24 What are the characteristics (eg investment in capital, level of exports, level of collaboration) of firms performing R&D and innovative firms? What is the performance (as measured by revenue, value-add, productivity) of firms performing R&D?	Low	High	H	H	H	We can answer this question well from current data sources, ie the LBD.
Data Q25 What are the factors driving firms' decisions on investment in R&D and who should perform it? What are the barriers to increasing investment in R&D?	Med	Med	M	H	H	The R&D Survey ask firms why they are investing in R&D, but does not ask about barriers to increasing investment in R&D.

COLLABORATION						
Data Q26 What is the extent, nature and quality of collaboration between researchers, either in New Zealand or offshore (eg, collaboration on joint R&D and innovation projects, co-authorship, co-invention in patents, contract R&D, exchanges, conference/event attendance)? Who is doing what with whom?	High	Med	M	M	L	This question can be partly answered by bibliometric data, patent office data, funding agency data, university and CRI reporting data. However, there is no agreed taxonomy of collaborations across the system which would identify the nature, type and extent of the collaboration. There is significant variability in what data funders collect on collaborations, eg some funders can determine the location of the collaborators, but few can identify how significant the collaboration is. Capturing data on sub-contracting and the research team is important for data on collaboration.
Data Q27 What is the extent, nature and quality of collaboration between researchers and end users (cut at project level), and between research institutions and end users?	High	Low	L	L	L	R&D Survey and BOS provide a useful high-level indicator of business-CRI/university collaboration, which we can track over time, but this does not tell us about the nature of the collaboration. No consistency between funders on information collected on end users, including information on co-funding. End user information is not linked to the LBD.
KNOWLEDGE EXCHANGE AND COMMERCIALISATION						
Data Q28 How much resource do universities and research institutes allocate to knowledge exchange activities?	Low	Low	L	L	L	Can be difficult to separate knowledge exchange activities from R&D activities, although institutes do have some staff dedicated to such activities. Knowledge exchange activities are often built into R&D contracts, but determining resources dedicated to knowledge exchange can be difficult. Information currently collected on knowledge exchange through reporting data is patchy. Contract reporting data on knowledge exchange activities is currently inconsistent across funding agencies. MBIE reporting data includes information on knowledge exchange activities, but we do not know how much resource is dedicated to those activities. CRI annual reports contain information on knowledge exchange activities, but not on resources allocated.

Data Q29 What are the citation rates of research outputs, including publications and patents, and who is doing the citing (cut by individuals, institutions, field of research, industry sector, domestic/international)?	High	Med	M	H	H	The coverage of bibliometric datasets is improving over time. The datasets allow us to analyse citations by individuals, institutions and domestic/international. Analyses by discipline and industry sector are more problematic. Patent datasets allow for analysis by individuals, institutions, discipline, industry sector and domestic/international.
Data Q30 What are the adoption rates or 'use' rates of research outputs and knowledge exchange activities of researchers and scientists, such as participation on panels, round tables and assessments, and engagement with policymakers in central and local government?	High	Med	M	M	M	Bibliometric datasets show the use of a publication by other researchers, but tend to show little of how a publication is used by end users, as end users seldom formally cite a publication. Adoption rates of particular innovations and technologies can be estimated through detailed analysis and working with industry groups or end user groups. Systemic data is difficult to capture. Knowledge exchange activities of researchers can be difficult to separate from R&D activities. Funding agencies' data on knowledge exchange activities is poor, although MBIE reporting does show whether the activities have taken place. Research information systems, including ORCID, are improving in capturing knowledge exchange activities such as membership on a panel.
Data Q31 How much direct revenue is generated from the commercialisation of New Zealand research (eg, patenting, licensing, spin-outs) and in what industry and field of research? Which researchers/ research providers were responsible?	High	Low	M	L	L	Pre-Seed Accelerator Fund (PSAF)1 data currently captures commercialisation activities of universities and CRIs, but quality and connectedness of data is poor - each institution has different ways of reporting the data due to differences in interpretations. Post contract follow-up by funding agencies is limited. Commercialisation data that MBIE collects through PSAF does not link back to researchers so it can be difficult to identify the links back to funding. In the absence of a system that links commercialisation activities and results to researchers, funding agencies would need post-contract follow-up, which is currently very limited.
Data Q32 What is the quality of commercialisation skills of New Zealand researchers, research providers and businesses?	Med	Low	L	L	L	Unable to answer this question. Partly a problem of determining how to measure the quality of commercialisation skills. Formal qualifications in commercialisation and entrepreneurship are growing, but they are limited as a measure of quality. The size of New Zealand's venture capital and angel investor markets could be a proxy.

Data Q33 What is the level and types of engagement/ communication between the science sector and the public?	Low	Low	M	L	L	MBIE's Science in Society survey seeks to answer this question but is fairly limited in coverage. The potential to use media and social sharing information has yet to be explored. Some funders ask for this information as part of knowledge exchange.
Data Q34 What are the levels of engagement between researchers and Māori?	Med	Low	L	L	L	Specific information on engagement with Māori is very limited. CRIs report on engagement with Māori in their annual reports, but there is little data. Vision Mātauranga profiling could give a sense of the expected engagement with Māori, but this profiling is not widespread across funding agencies.
INFRASTRUCTURE						
Data Q35 What is New Zealand's physical infrastructure for research and what is its financial value?	Med	Low	M	L	L	Individual institutions have financial information on their physical assets. MBIE's output expense for science infrastructure indicates the value of certain infrastructure assets. However, there is currently no register or map of research infrastructure akin to the EU map of research infrastructure. The R&D Survey contains capital expenditure on R&D which could form the basis of a national figure on R&D capital stock. The Annual Enterprise Survey captures capital stock and assets.
Data Q36 What is the demand for research infrastructure from different sectors?	Med	Low	L	L	L	Currently have difficulty answering this question, although individual institutions sometimes hold information on who is using particular infrastructure assets. Would need to generate information on who is using the existing infrastructure (after having determined what the infrastructure is).
COSTS						
Data Q37 What are the costs of doing research (ie inputs) in New Zealand?	Med	Low	L	L	L	No index currently available. Universities and CRIs have information on their researchers' salary costs, which form a large part of research costs.
Data Q38 What are the transaction costs in the science system, including governance costs of particular funding mechanisms?	Med	Low	L	L	L	Little information available. The occasional Surveys of New Zealand Scientists and Technologists have historically asked the following: "The percentage of total work time I spend on administration and compliance as opposed to research is: X%".

Appendix D: Links between enduring questions and data questions

DATA QUESTION	DATA LINK TO POLICY QUESTIONS	DESCRIPTION [note the policy questions identified are those with the strongest relationship to the data questions; there may be weaker links to policy questions which are not stated]
R&D funding and expenditure		
Data Q1 What research is being done in New Zealand? What is the breakdown of total R&D expenditure (*performing measure*) on an annual basis by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) performing institution; (vi) region where research is conducted?	Very high	Data on research performed in New Zealand is critical to all six policy questions, in particular: • policy question 1(a) on the effect of R&D performed in the public sector, 1(b) on the effects on firm performance of BERD, 1(d) on R&D and structural change • policy question 3(c) on the relationship between government funding and private sector funding for R&D • policy question 4 on the design of investments in science, such as 4(a) on absorptive capacity, 4(b) on specialisation versus broad base of investments, 4(i) impact of collaboration, 4(o) institutional arrangements and 4(p) the best use of research infrastructure • policy question 5 on government design for support of business R&D. Breakdowns of performing measures of R&D are important for analysing the flows between funding sources and performing entities. Investments in research are informed by relationships between the type of research performed and the performing entity, eg basic research may be less likely to be performed in the private sector, and applied research more likely to be performed in certain government entities. There may be relationships between the impacts of research and where the research is performed and the relationships may differ according to the type of research undertaken. Understanding these relationships is important to answering policy questions 1 and 2 on impacts, and investment in the system (questions 3, 4 and 5). The absorptive capacity of New Zealand business (policy question 5(a)) is best measured and understood by the amount and types of R&D they perform. R&D expenditure data is also used for analysis of R&D spillovers (3(b)).

Data Q2 How much does the government provide for R&D in total, and how is this split by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) research theme or government goal (including Vision Mātauranga); (vi) funder, funding programme, funding mechanism; (vii) performing institution, research provider type, researcher; (viii) region where research is conducted; (ix) benefitting region?	Very high	Data on government funding of R&D is critical to all six policy questions, in particular: • policy question 1(c) on the contribution to productivity of R&D financed by government, 1(d) on R&D and structural change, 1(f) and (g) on driving economic development of certain groups and regions • policy question 2, as most of this research is financed by government • policy question 3(a) on impacts and spillovers of government-funded R&D and 3(c) on the relationship between government funding and private sector funding of R&D • policy question 4(e) on the relationships between funding mechanisms, 4(g) the relationship between outcomes and particular funding mechanisms • policy question 5(c) on the impact of government R&D grants for firms. As the single largest funder of R&D, government funding needs to be broken down to inform policy settings and investment decisions. There may be relationships between the impacts of research, the type of research performed and the funding source. Understanding these relationships is important to answering policy questions 1 and 2 on impacts, and investment in the system (questions 3, 4 and 5). Determining the effectiveness of funding mechanisms requires comparison across funding mechanisms and this comparison requires taking into account the different types of research funded. Specific funds are sometimes established with specific socioeconomic objectives or horizons, and fields of research are important for organising the operations of funding agencies. Regional bodies are interested in research benefiting their regions and in the research performed in their areas. Funding agency data needs to be linked or enable comparisons to answer policy question 4(l) on duplication of funding and double-dipping. Data on government-funded research is also important for accountability and transparency purposes.
Data Q3 How much does the private sector (business, IROs and private non-profit) spend on R&D in total, and how is this split by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) performing institution and industry sector; (vi) region where research is conducted?	High	Data on private sector funding of R&D is important to all six policy questions, in particular policy question 5 on support for business R&D and policy question 1(c) on the contribution to productivity of R&D financed by the private sector. R&D funded by the private sector may be particularly important for answering policy question 1(d) on R&D and structural change. Government investment in R&D is informed by private sector funding of R&D (refer policy question 3(c)). There may be particular relationships between the types of research that the private sector funds and firm performance and other impacts. Hence, the breakdowns are important.

Data Q4 How much funding for R&D comes from overseas (other governments, international organisations, foreign firms) in total, and how is this split by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) performing institution; (vi) region where research is conducted?	Med	Data on overseas-funded R&D is important for policy question 3(d) on the contribution of foreign R&D to New Zealand, and policy question 1(c) on the contribution to productivity of R&D financed by overseas. There may be relationships between R&D funded by overseas sources and various impacts – either economic (policy question 1) or non-economic (policy question 2). Overseas-funded R&D may be performed in businesses and this may have an influence on the returns to R&D (refer policy question 5(a)).
Data Q5 How is government funding for R&D projects actually spent (including against its designated purpose)?	Low	Data on how government funding for R&D is spent informs policy questions 4(m) and 4(p) on infrastructure and facilities, and 4(o) on the effects of different institutional arrangements.
Data Q6 How much co-funding is leveraged for particular contracts, researchers, funding programmes, sectors, disciplines or research to particular socioeconomic objectives?	Med	Co-funding is a signal that research is expected to be useful to an end user. There may be relationships between co-funded research and various outcomes (policy questions 1 and 2) and these relationships may vary depending on the type of research. Co-funded research is a good indicator of connectedness (refer policy question 1(f) on the science and innovation system driving Maori and Pacific economic development and 1(g) on the system responding to regional economic development). Co-funding data is important for policy questions 4(d) on the responsiveness of institutions to end users and 4(o) on institutional arrangements.
Data Q7 What support ($) does government provide directly for hubs and precincts as well as incubators?	Low	This question informs policy question 5(g) on how clusters, hubs, precincts and incubators affect firm performance.
Data Q8 What are the funding sources of R&D performed in CRIs and universities by (i) government sources; (ii) non-government domestic sources; and (iii) overseas?	Med	This question can be answered if data questions 1, 2, 3 and 4 are complete. Therefore refer to those questions above.
Data Q9 How much are industry sectors required to spend on R&D into specific topics (due to legislative requirements)? How much levy funding is spent on R&D and in which industry?	Med	Data on levy-funded R&D informs policy question 1(c) on the contribution to productivity of R&D funded by levies. The data may also inform policy question 5(a) on the factors influencing the returns to R&D of New Zealand firms.
Data Q10 What is R&D as a percentage of revenue or added value in each industry (R&D intensity) and for each institution?	Med	This question can be answered through data question 1 and revenue figures available from the LBD for firms or institutions' annual reports. This data informs policy question 1 on industry-level economic outcomes, including 1(d) on R&D and structural change, where there may be relationships between R&D intensities, firm growth and industry growth. R&D intensities for research institutes inform policy question 4(o).

Data Q11 What are the characteristics of researchers and organisations applying for research funding? For each fund, who are the principal investigators and proposed contracting research organisations and what is the amount of funding sought? Science leader/principal investigator broken down by: (i) field of research; (ii) age; (iii) gender; (iv) ethnicity; (v) affiliated institution and type; (vi) qualification details; (vii) career stage. Research proposal data broken down by: (i) industry sector; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) research theme or government goal (including Vision Mātauranga); (vi) named researchers (proposed FTEs) and their IDs (Scopus, ORCID, NSN); (vii) proposed collaborations; (viii) length of project(s)?	Very High	The question informs policy question 4 and to a lesser extent policy question 6. Application data is especially important for: • policy question 4(i) on the impact of collaboration • policy question 4(l) on the effect of duplication in funding programmes and double-dipping • policy question 4(n) on funding success rates and bias. Application data is important in developing counterfactuals ie researchers not funded can be used as a control group against those who received funding, comparing each group's outputs. As researchers' demographics affect outputs, counterfactuals need to take into account demographic elements, so demographic information is required. Linking of application data can enable analysis of success rates and design of funding mechanisms eg we could determine whether a particular researcher cohort is being disadvantaged by particular funding mechanisms or across the board. Application data may also inform policy question 6, in particular 6(b) on strengths and gaps in research capability in New Zealand, and 6(g) on the development, retention and attraction of scientific talent. Application data by ethnicity and gender is also important to policy question 1(f) on the science and innovation system driving Māori and Pacific economic development, 2(b) on Vision Mātauranga and 2(c) on inclusiveness. Standardisation of application data would also reduce transaction costs for researchers.
R&D OUTPUTS		
Data Q12 What is the quantity and quality of R&D outputs of New Zealand-based researchers by: (i) field of research; (ii) socioeconomic objective; (iii) research type; (iv) horizons; (v) research theme or government goal (including Vision Mātauranga); (vi) funder, funding programme, funding mechanism; (vii) performing institution, research provider type; (viii) researcher characteristic; (ix) region where research is conducted?	Very High	Data on R&D outputs can inform all six policy questions. Policy questions 1 and 2 on impacts require tracking inputs, outputs and outcomes. Output data needs to be linked back to inputs for impact analysis, and this needs to be cut by types of research and associated with particular funds and researchers. Linking inputs and outputs requires certain classifications, hence the breakdown of the question. There may be particular relationships between types of outputs and certain outcomes (refer policy question 4g). Output data underpins many of the sub-questions of policy question 4, including 4(a) on research capacities and knowledge exchange, 4(c) contribution to global knowledge, 4(f) on the use of outputs, 4(h) on science quality and impact, 4(i) on the impact of collaboration, and 4(k) on the results from international connections and collaborations. Output data can also inform policy question 6, in particular 6(b) on the research strengths and gaps, 6(e) on the most effective mix of researchers, and 6(f) on the relationship between stability and outputs. In addition, output data is important for accountability and transparency purposes where government funding is involved.

PEOPLE AND SKILLS

Data Q13 What research capacity and capability exists in New Zealand? How many active researchers are there in New Zealand and how are they split by: (i) institution, industry sector, research provider type; (ii) career stage; (iii) field of research; (iv) qualification details (the level, place of study, and field of qualification); (v) funder, funding programme, funding mechanism; (vi) research project; (vii) age; (viii) gender; (ix) ethnicity?	Very High	This data informs policy question 6 and several sub-questions of policy question 4 (4(a) on research capacities and absorptive capacity and 4(b) on specialisation versus broad base of research capacities). Researcher demographic information is particularly important for: • policy question 2(b) on Vision Mātauranga • policy question 2(c) on inclusiveness • policy question 6(e) on how researcher attributes affect effectiveness (eg scientific output and quality) • policy question 6(f) on workforce stability and R&D outputs • policy question 6(g) on development, retention and attraction of talent • policy question 6(h) on the role of immigration in the science system.
Data Q14 What STEM skills exist in New Zealand and what are the trends over time?	High	This data informs policy question 6, in particular 6(g) on the development of scientific and research talent, and policy questions 4a on research capacities and 4(b) on specialisation versus broad base of research capacities.
Data Q15 What is the involvement and achievement of New Zealand students in STEM subjects? How is this changing over time? How is this split by: (i) age; (ii) gender; (iii) ethnicity; (iv) decile; (v) rural/urban split?	High	This data informs policy question 6, in particular 6(g) on the development of scientific and research talent. Also informs policy question 1(f) on the science and innovation system and Māori and Pacific economic development, and policy question 2(b) on Vision Mātauranga.
Data Q16 What are the post-study employment outcomes and career trajectories for Masters and PhD graduates and what does this look like by: (i) age; (ii) gender; (iii) ethnicity?	High	Data on post-study employment outcomes informs policy question 6(g) on the development, retention and attraction of scientific talent. It also indirectly informs policy questions 6(b) and 6(c) on research strengths, gaps and profiles, as employment outcomes and career trajectories may influence strengths and gaps. It also informs policy question 1(f) on the science and innovation system and Māori and Pacific economic development, policy question 2(b) on Vision Mātauranga and 2(c) on inclusiveness.

Data Q17 What is the mobility of New Zealand's researchers and what are the trends? What proportion and type of researcher (ANZSRC FOR) stay in New Zealand, go overseas, or return? What is the movement within the university and CRI sector and between public institutes and the private sector?	High	Data on the mobility of researchers can be used to examine spillovers (policy questions 3(a) and 3(b)). Data on research mobility can inform policy question 5(d) on barriers to knowledge exchange between public research organisations and firms, and policy question 5(f) on the diffusion of innovative practices. Research mobility can shed light on policy question 6(e) regarding the most effective mix of researchers, 6(f) on the relationship between workforce stability and R&D outputs and knowledge exchange, and 6(g) on the development and retention of talent. Relationships between research outputs, researcher mobility and firm performance could be carried out using the LBD and IDI. The relationships may differ by skills and field of research.
Data Q18 What is the level of managerial and entrepreneurial skills in New Zealand?	Med	Data informs policy question 4(j) on the appropriate balance of investment in knowledge creation versus other activities such as start-up support, and policy question 6(e) on how researcher attributes and skills influence effectiveness (scientific output, quality and measures of innovation).
Data Q19 Who are the top and emerging researchers?	High	This data feeds into policy questions 6(b) and 6(c) on research strengths and gaps and 6(e) on the most effective mix of researchers. It also informs policy question 6(g) on the development, retention and attraction of talent. This data may also inform policy question 4(g) on the association of outcomes with various outputs and inputs.
Data Q20 Who supervises research students?	Low	May inform policy question 4(o) on institutional arrangements and 6(a) on the relationship between teaching and research.
BUSINESS R&D AND INNOVATION		
Data Q21 What is the absorptive capacity of New Zealand business? What is BERD on an annual basis and what is the breakdown of BERD by: (i) firm size and distribution (eg, top 25, top 100 R&D spenders); (ii) industry; (iii) firm type; (iv) grant/non-grant recipients; (v) age of firm; (vi) length of R&D project?	High	Data on New Zealand business R&D is essential for policy question 1(b) on how BERD affects firm performance and 1(d) on R&D and structural change in the economy. BERD data is also required for 3(b) on industry spillovers. This data is essential for all of policy question 5 on government support for business R&D.
Data Q22 What are the innovation rates of New Zealand firms by: (i) firm size and by distribution; (ii) industry; (iii) firm type; (iv) grant/non-grant recipients; (v) age of firm? How much do firms spend on innovation?	High	This data is essential for policy question 1(b) on how innovative activity affects firm performance and 1(d) on how innovation drives structural change. Firm-level innovation data is important for policy question 5(a) on what factors influence the returns to innovation and 5(e) on what particular innovative activities lead to better firm performance.

Data Q23 What are the characteristics (eg investment in capital, level of exports, level of collaboration) of firms performing R&D and innovative firms? What is the performance (as measured by revenue, value-add, productivity) of firms performing R&D?	High	This data is essential for policy question 1(b) on how R&D and innovation affect performance and 1(d) on R&D/innovation and structural change in the economy. Also informs policy question 5.
Data Q24 How much does the Government spend on providing direct support to business innovation, excluding R&D?	Med	This data is most strongly linked to policy question 5(c) on the different R&D grants and support for innovation makes to firms and 5(e) on the relationship between particular innovative activities and firm performance.
Data Q25 What are the factors driving firms' decisions on investment in R&D and who should perform it? What are the barriers to increasing investment in R&D?	High	Data informs policy question 5(a) on the factors influencing the returns to R&D and 5(b) on the barriers to firm-level R&D.
COLLABORATION		
Data Q26 What is the extent, nature and quality of collaboration between researchers, either in New Zealand or offshore (eg, collaboration on joint R&D and innovation projects, co-authorship, co-invention in patents, contract R&D, exchanges, conference/event attendance)? Who is doing what with whom?	High	This question informs a variety of policy questions: • 1(f) on Māori and Pacific economic development • 4(a) specialisation versus broad base • 4(g) effect of collaborative funding arrangements on outcomes • 4(i) the impact of collaboration on quality and productivity of science • 4(k) the results from international collaboration • 4(o) the effect of different institutional arrangements • 4(p) co-ownership of research infrastructure • 6(e) the effectiveness of researchers and 'stars' and their networks • 6(g) the development, retention and attraction of talent.
Data Q27 What is the extent, nature and quality of collaboration between researchers and end users (cut at project level), and between research institutions and end users?	High	Collaboration between researchers and end users may influence impacts, so this data will inform policy questions 1 and 2. Data on researcher-end user collaboration also informs policy question 4(a) on research capacities and knowledge absorption (knowledge is transferred partly through researchers' links with end users), 4(i) on the impact of collaboration on quality of science, and 4(o) on the effects of different institutional arrangements. Data may also inform policy question 6(e) on the relationship between workforce stability and knowledge exchange.

KNOWLEDGE EXCHANGE AND COMMERCIALISATION		
Data Q28 How much do universities and CRIs spend on knowledge exchange activities?	Low	Primarily informs policy question 4(j) on the balance of investment in knowledge creation versus other activities such as knowledge exchange.
Data Q29 What are the citation rates of research outputs, including publications and patents, and who is doing the citing (cut by individuals, institutions, field of research, industry sector, domestic/international)?	High	Citation data provides indication of use, particularly use by other researchers and sometimes government agencies. Use data can be used for impact analysis (policy questions 1 and 2). Citation data informs policy questions: • 4(c) on the contribution of NZ R&D to the global pool of knowledge • 4(f) on the use of outputs • 4(g) on how different outputs may be associated with different outcomes • 4(h) the relationship between science quality and socioeconomic impact • 4(i) impact of collaboration on science quality • 4(k) results from international collaboration • 4(o) effects of different institutional arrangements • 6(e) most effective mix of researchers • 6(h) role of immigration in the science system.
Data Q30 What are the adoption rates or 'use' rates of research outputs and knowledge exchange activities of researchers and scientists, such as participation on panels, round tables and assessments, and engagement with policymakers in central and local government?	High	See above. Knowledge exchange activities may have an effect on the impact of R&D investments (and hence inform policy questions 1 and 2).

Data Q31 How much direct revenue is generated from the commercialisation of New Zealand research (eg, patenting, licensing, spin-outs) and in what industry and field of research? Which researchers/research providers were responsible?	High	This data may inform policy question 1(d), as commercialisation may have an effect on structural change in the economy over time. Commercialisation data also informs: • policy question 4(f) on the use of outputs • policy question 4(g) on the outcomes associated with particular outputs, types of research and funding mechanisms • policy question 4(j) on the balance of investment in knowledge creation versus other activities such as start-up support • policy question 5(d) on barriers to knowledge exchange between public research organisations and firms • policy question 5(g) on how clusters, hubs, precincts and incubators affect firm performance • policy question 6(e) on the effective mix of researchers.
Data Q32 What is the quality of commercialisation skills of New Zealand researchers, research providers and businesses?	Med	Data on commercialisation skills may inform policy questions 4(j) on the balance of investment in knowledge creation versus other activities such as commercialisation and 4(o) on institutional arrangements.
Data Q33 What is the level and types of engagement/communication between the science sector and the public?	Low	Policy question 6(i) on the effects of public engagement with science.
Data Q34 What are the levels of engagement between researchers and Māori?	Med	Primarily informs policy question 1(f) on the science and innovation system driving Māori and Pacific economic development, policy question 2(b) on how the science system contributes to Vision Mātauranga and policy question 2(c) on inclusiveness.
INFRASTRUCTURE		
Data Q35 What is New Zealand's physical infrastructure for research and what is its financial value?	Med	Policy questions 4(m) and 4(p) on research infrastructure and 4(d), 4(o) and 4(q).
Data Q36 What is the demand for research infrastructure from different sectors?	Med	Policy questions 4(m) and 4(p) on research infrastructure and 4(d), 4(o) and 4(q).
COSTS		
Data Q37 What are the costs of doing research (ie inputs) in New Zealand?	Med	Informs funding levels and can inform policy question 3 on marginal returns.
Data Q38 What are the transaction costs in the research system, including governance costs of particular funding mechanisms?	Med	Policy question 4(g) on the outcomes associated with particular funding mechanisms.

Appendix E: ORCID identifiers

ORCID is a non-profit organisation, operating as a charitable trust, incorporated in the USA on 5 August 2010. It is supported by a global community of organisational members, including research organisations, publishers, funders, professional associations and other stakeholders in the research ecosystem. The research community conceived and developed ORCID and set it up as a trust to ensure it remains platform-neutral. ORCID acts as a hub that connects with other researcher identification systems, publisher, funders, professional associations, repositories, and higher education bodies. A Board of Directors with wide stakeholder representation governs ORCID.

ORCID provides a persistent digital identifier that distinguishes each researcher. ORCID's key goal is to unambiguously identify researchers and provide tools to automate the connection between researchers and their creative works. ORCID supports automated linkages between the researcher and the researcher's activities and outputs through integration in key research workflows such as manuscript and grant submission. For instance, ORCID is integrated with publishers such as Thomson Reuters and Elsevier and research data registries such as DataCite. ORCID has worked with Crossref and DataCite to ensure an individual's ORCID record is automatically updated when their work is published.

ORCID could be used as a single location for academic activity and, when incorporated into existing data sources, can also be used as a digital 'key' to aggregate information from these data sources, for example, publisher catalogues, institutional archives, data archives, HR records and funding management systems. These aggregations of outputs can then be accurately attributed to the individuals involved and certain types of impact more easily measured.

A number of unique author identification systems are available, such as International Standard Name Identifier (ISNI) or the Virtual International Authority File (VIAF). For researchers, three main unique author identifiers exist: ORCiD, ResearcherID and Scopus author identifiers. However, only ORCID is a non-proprietary, platform-neutral identifier which links a researcher and affiliation with other identification schemes. ResearcherID integrates with the Web of Science platform and Scopus IDs integrate with Scopus and ScienceDirect databases. ResearcherID and Scopus records can be integrated with ORCID records. Unlike other identifiers, ORCID is not limited by discipline or by geographic region. It can be used by all researchers, not just those in the 'hard' sciences.

ORCID was launched in October 2012 and in three years has served more than 1.7 million IDs. A large number of institutions, including publishers, libraries, societies and research providers, are integrating with ORCID.[42] Member organisations connecting to the ORCID registry include Altmetric, Australian National Data Service, British Library, CrossRef, figshare, the US Food and Drug Administration, Elsevier, Harvard University, Nature Publishing Group, Oxford University Press, ProQuest, Researchfish, Springer Science, Symplectic, Thomson Reuters, University of Cambridge, UK National Institute for Health Research, US National Institutes of Health and the Wellcome Trust.

Some funders internationally are now requiring the use of ORCID identifiers:

- Swedish Research Council from 2015
- Wellcome Trust from 2015
- UK National Institute for Health Research from 2015
- Science Foundation Ireland from 2016
- Austrian Science Fund from 2016.

42 http://orcid.org/organizations/integrators/current.

The UK Metrics Tide report[43] recommended that ORCID IDs be mandatory for all researchers in the next Research Excellence Framework (REF) assessment. The report also recommended that UK funders and higher education institutions use ORCID for grant applications, management and reporting platforms.

ORCID delivers benefits to all parts of the research system, from the researcher to the research institute and onto the funder and policy agencies. The broad benefits of using ORCID identifiers include:

› The ability to disambiguate researchers through an online identifier that links to a researcher's works (publications, datasets etc), links to other researcher identification systems, and is retained regardless of a researcher's institutional affiliation
› Enabling researchers to interact with multiple institutions, publishers and funders in New Zealand and around the world through the use of a common identifier
› Simplification and automation of data entry processes, reducing the administrative burden on researchers, research institutions and funding agencies
› Addressing duplication of effort and enabling the reuse of data for multiple purposes, both within an organisation and across organisations through automation of processes and data exchanges
› Improving data quality (accuracy, completeness, consistency, validity etc) through automated data extraction, harvesting and testing across systems and organisations
› Reducing the workload for researchers, research institutions and funding agencies in the long term, improving the overall efficiency of the research and science system as a whole
› Providing the necessary infrastructure to integrate data and facilitate timely and efficient data collection for the management of a research institution, and more broadly for monitoring the performance of New Zealand's research and science system
› Enhancing the online presence and exposure of New Zealand researchers and their research activities and outputs to the global market, industry partners, international collaborators, and students aspiring to study in New Zealand.

43 Wilsdon et al (2015).

References

Alston, J.M., M.A. Andersen, J.S. James and P.G. Pardey (2010), *Persistence Pays: U.S. agricultural productivity growth and the benefits from public R&D spending*, Springer: New York.

Australian Bureau of Statistics (2015), *Research and Experimental Development, Businesses, Australia, 2013–14: Explanatory Notes*, Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/8104.0Explanatory%20Notes12013-14?OpenDocument.

Baghai, M., S. Coley and D. White (2000), T*he Alchemy of Growth: Practical insights for building the enduring enterprise*, Perseus: New York.

Crawford, R., R. Fabling, A. Grimes and N. Bonner (2007), *National R&D and Patenting: Is New Zealand an Outlier*? New Zealand Economic Papers, 41 (1), pp. 69–90.

Curry, A. and A. Hodgson (2008), 'Seeing in Multiple Horizons: Connecting futures to strategy', *Journal of Future Studies 13* (1) pp. 1–20.

Davis, N. (2006), *Business R&D, Innovation and Economic Growth: An evidence-based synthesis of the policy issues*, Ministry of Economic Development: Wellington.

Department of Health and Human Services US (2015), *Biomedical Research and Development Price Index (BRDPI): Fiscal Year 2014 update and projections for FY 2015-FY2020*, National Institutes of Health, Retrieved from https://officeofbudget.od.nih.gov/pdfs/FY16/BRDPI_Proj_Jan_2015.pdf.

Edler, J., P. Cunningham, A. Gok and P. Shapira (2013), *Impacts of Innovation Policy: Synthesis and Conclusions*, Nesta Working Paper No. 13/21, Retrieved from http://www.nesta.org.uk/sites/default/files/impacts_of_innovation_policy_synthesis_and_conclusion_final.pdf.

Guellec, D. and B. van Pottelsberghe de la Potterie (2001), *R&D and Productivity Growth: Panel Data Analysis of 16 OECD Countries*, OECD Science, Technology and Industry Working Papers, 2001/03.

Hall, B.H. and A.B. Jaffe (2012), *Measuring Science, Technology, and Innovation: A review*, A report prepared for the Panel on Developing Science, Technology, and Innovation Indicators for the Future, National Academies of Science. Retrieved from https://eml.berkeley.edu/~bhhall/papers/Hall-Jaffe%20HJ12_indicators_final.pdf.

Hall, B.H., J. Mairesse and P. Mohnen (2009), *Measuring the Returns to R&D*, NBER Working Paper No. 15622, Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1530078.

Hall, J. and G. Scobie (2006), *The Role of R&D in Productivity Growth: The case of agriculture in New Zealand: 1927 to 2001*, Working Paper 06/01, The Treasury: Wellington.

Jaffe, A. and T. Le (2015), *The impact of R&D subsidy on innovation: a study of New Zealand firms*, Motu Working Paper 15-08, Retrieved from http://motu-www.motu.org.nz/wpapers/15_08.pdf.

Johnson, R.J., W.A. Razzak and S. Stillman (2007), 'Has New Zealand Benefited from its Investments in Research and Development?', *Applied Economics 39*, pp. 2425-2440.

Lundvall, B-A. (ed.) (1992), *National Innovation Systems: Towards a theory of innovation and interactive learning*, Pinter: London.

Madsen, J.B. (2010), 'The Anatomy of Growth in the OECD since 1870,' J*ournal of Monetary Economics 57* (6), pp. 753–767.

Ministry of Business, Innovation and Employment (2015), N*ational Statement of Science Investment 2015–2025*, Retrieved from http://www.mbie.govt.nz/info-services/science-innovation/pdf-library/NSSI%20Final%20Document%202015.pdf.

Ministry of Business, Innovation and Employment and the Ministry of Education (2014), *Tertiary Education Strategy 2014–2019*, Retrieved from http://www.education.govt.nz/assets/Documents/Further-education/Tertiary-Education-Strategy.pdf.

Ministry of Economic Development (2011), *Evaluation of the impacts of cross-vote government assistance on firm performance stage 2: Impacts of direct financial support for R&D*. Retrieved from http://www.mbie.govt.nz/publications-research/publications/evaluation-of-government-programmes/evaluation-of-cross-vote-government-assistance-on-firm-performance-april-2011.pdf.

Nelson, R.R. (ed.) (1993), *National Innovation Systems: A comparative analysis*, Oxford University Press: New York/Oxford.

New Zealand Institute of Economic Research (2015), *Impact of the Cawthron Institute: Economic contribution to Nelson and New Zealand*, NZIER report to the Cawthron Institute, NZIER: Wellington, Retrieved from http://www.cawthron.org.nz/media_new/publications/pdf/2015_10/NZIER_Economic_Impact_Report_Oct_2015.pdf.

OECD (1997), *National Innovation Systems*, OECD: Paris, Retrieved from http://www.oecd.org/science/inno/2101733.pdf.

OECD (2001), *Measuring Expenditure on Health-related R&D*, OECD: Paris, Retrieved from http://www.oecdbookshop.org/get-it.php?REF=5LMQCR2K9H22&TYPE=browse.

OECD (2002), *Glossary of Key Terms in Evaluation and Results-Based Management*, Paris: OECD, Retrieved from http://www.oecd.org/development/peer-reviews/2754804.pdf.

OECD (2005), *The Impact of R&D and Human Capital on Productivity Growth*, Paris: OECD, DSTI/EAS/IND/SWP(2005) 9.

OECD (2010), *The OECD Innovation Strategy: Getting a head start on tomorrow*, Paris, OECD.

OECD (2013), *Linking Science to Innovation, Economic Growth and Welfare: Scoping Paper*¸ DSTI/STP/RIHR(2013)4/REV1.

OECD (2015), *Frascati Manual 2015: Guidelines for collecting and reporting data on research and experimental development*, The Measurement of Scientific, Technological and Innovation Activities, Paris: OECD.

Pells, S. (2014), H*ow can we lift BERD*?, Ministry of Business, Innovation and Employment Occasional Paper, Wellington.

Ridley, R. (2014), *PreSeed Accelerator Fund Outcomes: 10 Year Review*, Ridley Insight Ltd, Retrieved from: https://www.kiwinet.org.nz/files/Investment/Historic-PreSeed-Report.pdf.

Salter, A.J. and B.R. Martin (2000), 'The Economic Benefits of Publicly Funded Basic Research: A critical review', *Research Policy 30* (2001), pp. 509-532.

Wallace, M.L. and I. Rafols (2015), *Research Portfolio Analysis in Science Policy: Moving from financial returns to societal benefits*, Science Policy Research Unit (SPRU) Working Paper Series SWPS 2015-10, Retrieved from https://www.sussex.ac.uk/webteam/gateway/file.php?name=2015-10-swps-wallace-rafols.pdf&site=25.

Westmore, B. (2013), *R&D, Patenting and Growth: The role of public policy,* OECD Economics Department Working Papers No. 1047, Retrieved from http://dx.doi.org/10.1787/5k46h2rfb4f3-en.

Wieser, R. (2005), 'Research and Development Productivity and Spillovers: Empirical evidence at the firm level,' *Journal of Economic Surveys 19* (4), pp. 587–621.

Wilsdon, J. et al. (2015), *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*, Retrieved from http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/The,Metric,Tide/2015_metric_tide.pdf DOI: 10.13140/RG.2.1.4929.1363.

MB13609 SEP16

ISBN 978-0-947524-25-8 (Online)