eResearch 2020

National Research Data Programme

The Case for Research Data

About eR2020

eResearch 2020 is a future-oriented national programme that develops strategy with thought leaders across the research sector and aims to assemble a comprehensive vision of researcher needs and essential skills over the coming decade. eResearch 2020 is led by NeSI, REANNZ and NZGL as co-patrons, together taking a combined approach to facilitating national discussions. Fundamentally cross-institutional and cross-discipline, eResearch 2020 brings researchers together to focus on particular themes, be it on research sector cloud strategies; skills gaps; institutional governance, research capabilities, or the infrastructure needs of the National Science Challenges and the Centres of Research Excellence.

The proposed National Research Data Programme has been developed through eResearch 2020 national workshops in April and June 2014, and in July 2015, as well as through the eR2020 Interview Series (www.eresearch2020.org.nz) and the 2015 discussion document "eResearch Challenges in New Zealand".

eResearch 2020

VISIT OUR NEW WEBSITE:
eresearch2020.org.nz

REANNZ

NeSI

NZGL
NEW ZEALAND GENOMICS LIMITED

New Zealand invests over $1.4bn in publicly funded research every year, yet every year we fail to realise the full value of our investment. This is because we don't connect or coordinate our investments in research data management; we don't incentivise our researchers to train in digital research methods, and we often struggle to effectively manage the data our researchers produce. Most tellingly, we fail to connect our researchers to decisions in our economy and society that would benefit from better data management. eR2020 proposes that coordinated, sector-wide action on digital research and data can produce both direct and network benefits, and that these benefits can lift the return on the $1.4bn New Zealand invests publicly funded research each year.

We think the timing is right for an integrated national approach to research data. We can see new research leaders emerging, with new skills, in what is an ageing research sector workforce. We also observe that there's wider economic and social demand for data enablement, higher digital standards for research publication, and that technology costs are falling.

Overall, we think stronger active links for research data to business R&D and public sector decision-making will contribute towards more impactful research, and may lift digital enablement in our society and the economy of New Zealand. This, along with strong support for researcher data skills and better data management in research institutions, can begin to recoup some of the estimated $392m pa of unrealised value of data in the research sector.

Contents

Executive Summary

This eResearch 2020 Position Paper proposes an economic and strategic case for a programme of change across the research sector aimed at transforming New Zealand researchers' abilities to work with and create value from data. Overall, eResearch 2020 analysis suggests that an investment of approximately NZ$50m - NZ$70m over a 5-year period in a National Research Data Programme (NRDP) could stimulate up to NZ$392m per annum in unrealised value within the research sector, and make a permanent contribution to GDP over a 15-year period.

Underpinning the proposed Programme is the strategic objective of increasing the quality and impact of New Zealand research output, along with lifting New Zealand researcher participation in an increasingly digital and data-intensive global research environment. The NRDP role would be to identify, evaluate, and execute on those intermediate interventions and activities that cumulatively contribute to this objective. As is the case with a transformative programme of change, we have not scoped all of the different projects or activities that might be required to achieve the goal, rather we are seeking to identify insight into appropriate opportunities for coordinated action at a national level that will have the greatest bearing on the whole Programme objective.

Thanks to our work across the sector developing the 2015 discussion document, eResearch Challenges in New Zealand, we've learned that a national approach needs to focus on developing excellence in the first steps of the research process, where research methods are employed to collect or create data. Our observations also suggest that structured, professionally managed "data bridges" that actively connect researchers to the users of research data will enable both greater research impact and better, data-informed decision making across multiple sectors in our economy and society.

First Use of New Data

We want the NRDP to focus on the First Use of New Data. We think that focusing on improving the impact of the first use of newly created or collected data will make the biggest contribution towards the objective of more impactful research, as our researchers and institutions are already resourced to deliver these research outputs. A focus on the first use of data might also influence excellence in methods, standards, tools and researcher behaviours at the beginning of the research process and data life cycle, which we think will have flow on effects towards quality and impact across the research system.

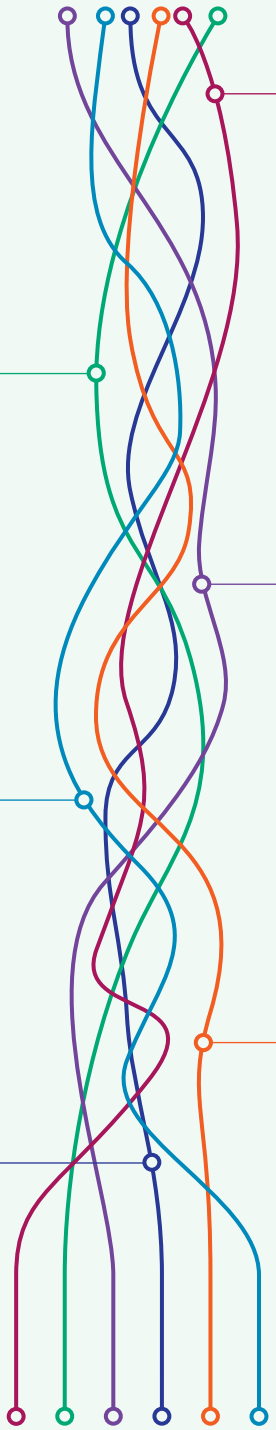
Active Data Bridges

A successful Programme needs to create and expand Active Data Bridges between research, industry, and government. We think that a focus on digital bridges to industry and government that allow data to influence and empower decision making will create economic outcomes and social benefits for New Zealand. Well-managed and resourced data bridges for active data (i.e. non-historical data) will better enable data-empowered decision making and better connect our researchers with their digital customers and the users of research data. What's more, better connections to the customers of research often offers researchers insights that lead to more impactful, applied research outcomes.

FIGURE 1: PROGRAMME OUTCOMES. THE HIGH LEVEL INITIATIVES WITHIN THE PROPOSED PROGRAMME MIGHT INCLUDE.

Programme Outcomes.

The high level initiatives within the proposed Programme might include:



A national meta-data catalogue that can implement open data policies, improve the impact and efficiency of research by lowering barriers to sharing, and co-locate data for major cross-sector initiatives and collaborations such as the National Science Challenges (NSCs).

Development of active data bridges from research to government (central and local) and industry that support data-empowered decision making, productivity and innovation.

A Programme implementation capability resourced and empowered to implement the major Programme initiatives in collaboration with key partners across the wider research sector.

Aligned national and institutional policies and funding mechanisms which actively encourage institutions and researchers to develop data intensive research capabilities and programmes. In a digitally enabled sector, alignment needs to include people, methods, infrastructure and support, both within institutions and in collaboration with other institutions nationally and internationally.

A comprehensive professional development programme rolled out through the sector aimed at lifting New Zealand researchers' baseline skills with data and digital methods, in line with sector agreed standards, international norms, and researcher expectations and supported by dedicated, discipline-led technical standards groups.

Operational support for data management, including skills, storage, and national collections where appropriate and particularly within research institutes with statutory data responsibilities.

Key observations of New Zealand research data and digital capabilities that contribute to our analysis:

Observation

We already invest NZ$1.4bn each year in data creation and data collection, as an activity within our fully-funded research system – in one estimation, 46% of researchers' time is spent collecting and analysing data.

The most costly activities in research data management are the costs of acquiring research data, followed closely by the costs of sharing research data.

Increasing the volume of data typically reduces the per unit costs of data management.

Research data management and digital methods make a contribution to the functioning of our buildings, our roads and ports, our health system and our industries.

The expectations of top international journals and peer reviewers are changing – research data is expected to be accessible, and the digital methods underpinning research findings must be reproducible.

FIGURE 2. KEY OBSERVATIONS

Impact

This suggests that improving digital research methods, or enabling more ambitious research to be conducted by larger teams will be the most direct approach to lifting the impact of this annual investment.

Focusing on the first use of new data is where the biggest long-term gains are to be made.

Economies of scale are accessible in data, if we act at a national level.

Creating better digital engagement between the research sector and the users of research data can have direct impact on New Zealand society as a whole.

New Zealand researchers need access to the expertise and infrastructure to ensure they are able to excel against these new expectations.

The intermediate outcomes and key projects that would make up the Programme have been sourced from the research community through interviews and workshops, and represent the best demand-side thinking we have available at this time. This work has been augmented with macro-economic and cost/benefit analysis from experiences in Europe, North America, and Australia.

Developing the Strategic & Economic Argument

As we are proposing a programme of work, this paper sets out the possible strategic and economic case for change, and outlines some of the functions a successful NRDP might deliver. Once established, it would be the role of the Programme to develop the specific projects, and implement the infrastructure, skills, and foundational capabilities required. Along with enabling infrastructure and services, the shift towards digital research methods includes professional development, common standards and meta-data capabilities, and implementation of appropriate IT platforms and international linkages, along with appropriate policy adjustments.

In a strategic sense, we think this opportunity for whole sector development is available now, due the ageing demographics of the research sector workforce and the emergence of a new generation of leaders in research who are keen to engage with digital opportunities. Furthermore, the change in cultural attitudes towards data sharing and widespread adoption of digital technologies, along with the falling costs of data-related technology and services, suggest that coordinated action is likely to be affordable, well-received, and timely in terms of demand. Beyond this, we see development of digital capability in the research sector as an opportunity to significantly expand our national research horizons. Feedback from the sector suggests a broader digital capability will also allow us to do much more with the research funding and resources we already have.

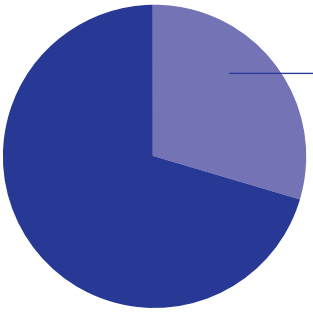
From an economic perspective, we have sought to apply a tested international methodology for evaluating research data value to the New Zealand sector, and asked independent economists at PwC to adapt these methods for the New Zealand economy.

Once established, the NRDP would have the opportunity to extend this analysis, however our review suggests that the value of research data created each year under our fully-funded research system could be as much as NZD589m, but that we have no mechanism for sustaining that stock of knowledge, therefore much of that value is lost from year to year.

TOP: FIGURE 3. ECONOMIC IMPACT

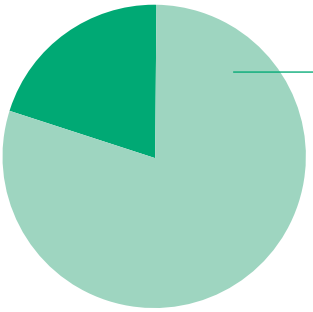
BOTTOM: FIGURE 4. THE VALUE OF BETTER RESEARCH DATA MANAGEMENT IN NZ

ECONOMIC IMPACT UPPER BOUND FIGURES ONLY



$589m pa
Proportion of Govt. total research spend invested per annum in collecting, creating, analysing and managing research data.

$1.4bn pa
Total Govt. pa spend on publicly funded research (2014).

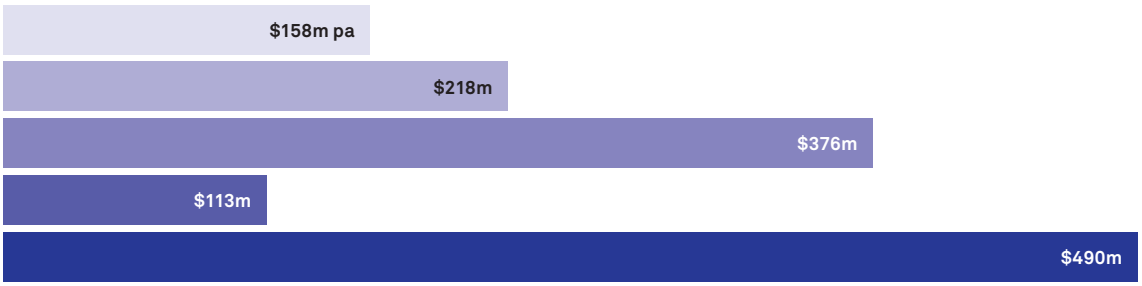


$392m
80% unrealised / lost value in publicly funded research per annum, due to sub-optimal management of research data.

$490m
Potential return (NPV) on 1 year of publicly funded research data activity (ROI over 20 years).

THE VALUE OF BETTER RESEARCH DATA MANAGEMENT IN NZ UPPER BOUND FIGURES ONLY

- Direct time-efficiency/cost savings to researchers from working with better managed data
- Localised benefits of greater efficiency, reinvested in the research system
- Combined Impact of cost savings and greater research efficiency
- Value of additional use of 1 year's data if managed in repositories over a 10-year period
- Total additional impact from better research data management



Treat Data as a Strategic Asset:

FIGURE 5. DATA AS A STRATEGIC ASSET

"We recommend that government, with business and communities, develop strategies to ensure New Zealand can unlock our data assets and drive economic and social value."

The Opportunity of Research Data

Technology is remaking the world around us, and the opportunities for human discovery are as vast as they are thrilling. ICT intensive research, or "eResearch" has found the Higgs-Boson particle, identified earth-like planets across the galaxy, detected gravity waves, created genomics as a powerful force in our lives (along with a few other "omics"), and will soon be driving our cars! Some of the world's largest companies have been built on data and computing power – Amazon and Google are becoming the old guard in this respect and yet even they are rushing into new areas of digital research. Few New Zealand companies need to be sold on the benefits of digital and data analytics, even though the investment required by digital initiatives is substantial. Unfortunately, the story is less glossy for the New Zealand research sector; where we still have considerable opportunity to grow our participation in digital globalisation and data-intensive research.

Over the last 25 years, our universities have become some of the largest and most advanced ICT organisations in New Zealand. The Government has invested heavily in ICT and digital infrastructure at a national level, as well as in eResearch with NeSI, NZGL, REANNZ and the Synchrotron. Our research institutions house some of the highest tech equipment and most advanced skills to be found in the country. Some New Zealand researchers regularly punch well above their weight in global terms, yet overall we have seen neither the emergence of any broad digital capability in our research sector, nor the skills diffusion into society and the economy we had hoped this would engender. Our high tech equipment is not connected, the output not shared. Our skills are in silos, our research projects often disconnected, ad hoc collaborations such as the National Science Challenges (NSCs) are launched lacking the digital fundamentals. We need to reach a point where new national initiatives in research can be born digital, with a strong data foundation and the security that the knowledge created will add value into the long term.

Flows of physical goods and finance were the hallmarks of the 20th-century global economy, but today those flows have flattened or declined. Twenty-first-century globalisation is increasingly defined by flows of data and information.

DIGITAL GLOBALISATION: THE NEW ERA OF GLOBAL FLOWS: MCKINSEY GLOBAL INSTITUTE, MARCH 2016

In global terms, the ability of humanity to create, collect, store and share data has never been greater. From a research perspective, the scale of global challenges and the scope of computation and data available to tackle them are breathtaking. Digitised literature, data from public services, or data generated by connected sensors can now be collected, stored, and analysed like never before. Mobile technology and the rise of social media offer unprecedented opportunities for researchers to contribute in new ways and with immediate feedback. With the right skills and the right tools, we could have the opportunity to understand and manage our environment, our primary industries, and our urban systems with greater detail and precision than ever before.

In general, New Zealand has been a late adopter of technology and change in the research world, for the very good reason that while adopting early is risky and costly, these risks and costs fall over time. National data collections, research discipline data repositories, sensor networks and data intensive discovery have all emerged over the last 20 years in the international domain; however in the last 5 years these technologies have matured rapidly. Since 2010, the capabilities of data infrastructure and services (research and otherwise) have exploded even as the costs of the necessary underlying technology and systems have plummeted. At the same time, a new generation of research leaders are emerging who are, generally speaking, enthusiastic to embrace the scale of technological change that's occurring. Getting better at digital methods and data-intensive skills at a research system-wide level is a complex challenge, however now is the time to get on with it.

Notable Exceptions: NZ researchers invented R, and Weka, two key platforms with broad scale adoption internationally – so we innovate, though we don't diffuse and absorb our own innovations. 23+ years since R was developed, it is a key language at the heart of a global revolution, including one which is taking hold in our government agencies, and has been a mainstay of international research for a couple of decades.

NICK JONES, DIRECTOR, NESI

The Change Occurring in Research

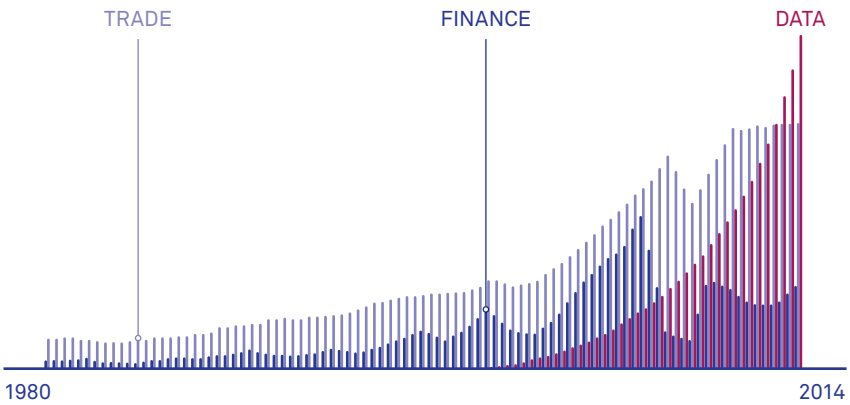
At the most fundamental level, the role of "data" in research is changing at pace. Sir Peter Gluckman, Chief Science Advisor to the Prime Minister, suggests data has been transformed from an output of hypothesis testing into a starting point for hypothesis generation. The advent of the "Cloud" – the infrastructure for managing, transporting and analysing huge and diverse flows of data - is a phenomenon that is changing our economy and society, and enabling major discovery and innovation. "Data-intensive Discovery" is rapidly altering the fundamentals of science systems around the world. Unprecedented collaboration (papers with over 1,000 authors), new models for funding hypothesis-free research, and even more rapid and widespread sharing of research findings all contribute to new expectations of research and researchers.

As it stands, New Zealand is somewhat underprepared for this evolution of the role of data, for three reasons. Partly this is a generational issue, the leaders of our research sector are from the baby-boomer generation and many are nearing retirement age – the contribution of their experience is in their research outputs, rather than in new technologies and methods for conducting research. Partly this is a resourcing issue; just how much time do we expect our research institutions to spend thinking about digital research and data, and have we made our expectations clear? And partly this is a confidence issue; our researchers are clustered in relatively small geographical groups, they don't have broad access to baseline professional development in this area, and they aren't sure how "big" they could be thinking. While being a late adopter has certain cost and risk advantages, it will be important we take a coordinated approach to these challenges, as becoming a digital laggard in research may run an entirely new set of risks.

The Rise of Data Analytics

As the fundamentals of mobile technology shift, the societal and cultural expectation will become that data, be it climate data, health data, language or cultural data, will be available. Underlying data collections, such as genomic records, environmental observations, traffic flows, and biological integrity databases, become the "data infrastructures" that will support decision-making in government, in public services such as health and social development, along with productivity growth, innovation and marketing. The recently released 30 Year New Zealand Infrastructure Plan makes this point explicit, highlighting both the real-time nature of data and the impact common standards and open data can have on infrastructure management. While these are broad-brush examples of societal reliance on data, the digital

standards and tools that underpin these data infrastructures, along with a large percentage of the data that feed these systems, are typically generated and maintained within the research sector.



45 times growth in data flows 2005 - 2014

FIGURE 6. GLOBAL FLOWS OF TRADE AND FINANCE ARE FLATTENING, WHILE DATA FLOWS ARE SOARING

The current MBIE review of the Telecommunications Act 2001, "Regulating Communications for the Future," points out that convergence is reshaping our telecommunications and broadcasting markets, that the Internet of Things will play an increasing important role, and that the Internet is eroding national boundaries for content and services. McKinsey, a consultancy, recently noted that, as global data flows have increased at stunning rates, digital infrastructure has become as important as transportation infrastructure in growing national economies, however they point out that New Zealand participation in the digital globalisation is lagging up to 75% behind the developed world. Our research sector can help NZ catch up, yet for digital success in the research sector, certain capabilities – especially those that build foundations for other key processes and activities – will be more important than others. Foremost among them are IT platforms that can overcome fragmentation, enable effective research, and cope with the peak demands of research equipment and the inter-relational interdependence among researchers (often working from diverse locations) required for high impact research. The typical IT platforms of our research institutions are organised to prioritise their larger franchise corporate IT and student-focused consumer needs, making it difficult for them to recognise or service the small to medium enterprise nature of research and research data. Change is upon us; McKinsey also reckon over 900M individuals are participating in digital globalisation directly, 80% of tech start-ups are born global, and that the impact of data flows on world GDP growth over the last 10 years has been larger than the global trade in goods.

Once the IT platforms begin to address research data needs, the focus shifts to the capability of our researchers to engage with their "customers" digitally, be they the scholarly customers of their collaborators and peers in the research world, or the customers of research outputs in central or local government, or in industry and society. Impactful research requires researchers to be engaged with each other for collaboration, and connected to the users of research. In a digital world, this requires the digital capability to collaborate, and to translate findings into data-empowered decision making.

Building Foundational Capabilities

The technological and organisational changes required to shift our research sector towards digital methods and data-intensive research are complex and challenging. For firms, experts highlight "Mind-Sets" and "Technology Systems" as the two fundamental elements for successful digital enterprise. We think these two major elements translate well in developing focus for New Zealand researcher efforts such as our National Science Challenges, and designing a National Research Data Programme:

Mind-Sets & Skills. being digital is about using data to make better and faster research decisions; expanding research goals to bigger, more impactful work across larger, more diverse teams, and developing much more iterative and rapid ways of doing things. Thinking in this way shouldn't be limited to the top scientists of leading disciplines, it should incorporate the broad scope of researchers' work, including collaborating across disciplines, institutions, and with industry and the public service to extend necessary capabilities. A digital mind-set builds environments to encourage the generation of new ideas, and values the digital and data assets as both metrics of success and keys to supporting decision-making agility.

Technology Systems & Data. in the context of ICT, digital research is focused on creating a two-speed environment that decouples legacy or consumer systems (which support corporate and commodity functions and often run at a slower pace), from those that overcome research fragmentations, and support complex, sometimes peak performance interactions (often in a highly interdependent way). A key feature of digital research IT is the commitment to connecting devices, instruments, objects, and people with the meta-data architecture, standards, storage, and analysis that can enable interoperable, systems-level coordination in research.

Data Dynamics

Data is a useful umbrella term across many aspects of our lives, however not all data are equal. In research, data comes in many shapes and forms. Unsurprisingly, different types of data are required for different research disciplines, each with a different approach to meta-data. Data can be created through computation, collected iteratively over a time, streamed constantly from a sensor or network of sensors, or developed from observations. Research data is not just generated by researchers; central and local government, environmental agencies, and the public – clinicians, teachers, and farmers – all develop research data. Our New Zealand research system is actually very good at creating and collecting data; however up until quite recently technological constraints on storage and connectivity meant that most of that data could not be effectively retained or shared. In an increasingly digital society and economy, we find ourselves relatively inexperienced at managing our data, and under-resourced if we're to sustain this stock of knowledge.

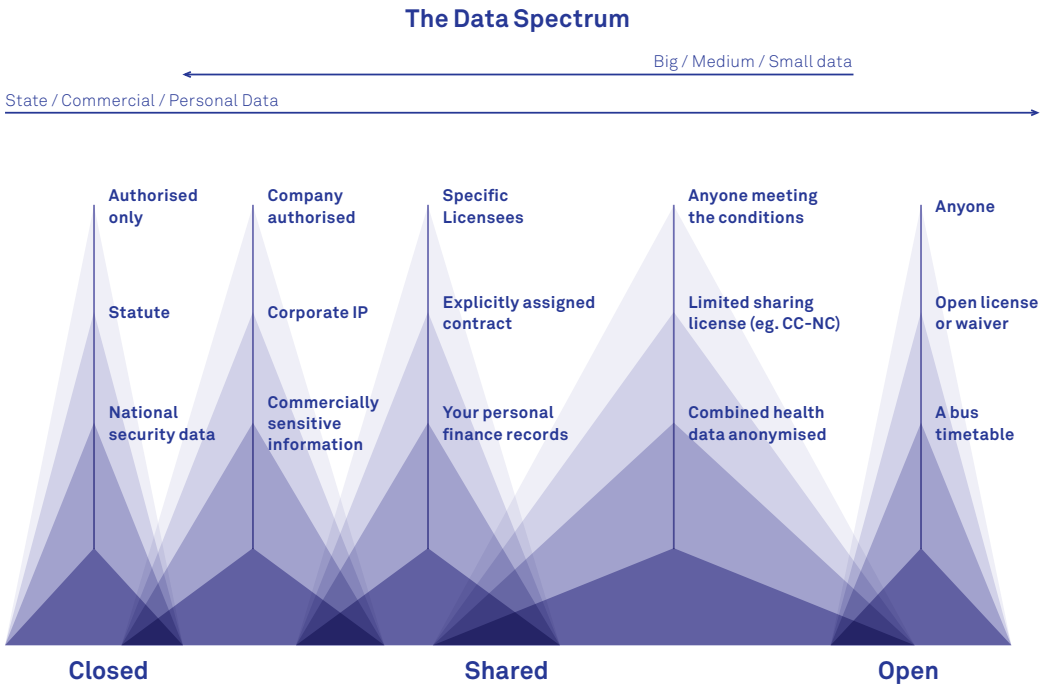


FIGURE 7. THE DATA SPECTRUM. (HTTP://THEODI.ORG/DATA-SPECTRUM - THE OPEN DATA INSTITUTE)

Open Data and Data Discoverability

Open Data means that the existence of particular data can be discovered, and access to that data can be negotiated. It does not mean that access will be free; data is expensive to collect and valuable to employ. It also does not mean that all data will be made available – there are many reasons why data might be confidential or private, and knowing it exists doesn't automatically convey a right to see it. The London-based Open Data Institute uses a helpful diagram (Figure 7) to differentiate data by access.

In principle (and possibly in law), publicly funded research data is supposed to be Open. New Zealand research institutions will need some help to achieve this, however it's a goal worth striving for. A well-functioning, implemented open research data policy not only increases the impact and efficiency of research, we think it will also encourage trusted data bridges to form between research, industry and government which have potential to transfer knowledge as effectively as journal publishing and possibly a lot faster.

On the Data Spectrum at Figure 7, some data is secret - its very existence is restricted knowledge. There will always be cases where this is needful, and the technical solutions to manage this are well-understood, provided the data management capability exists in the sector in the first place. As we don't yet have a strong national research data capability, our institutions are forced towards imprecise or broad-brush secrecy, making knowledge unavailable for sharing for longer than might actually be necessary.

Active Data versus Passive Data

We think it's important to distinguish between active and passive data in research. Active Data is generated through research activity every year – it is data our researchers are working with and developing outcomes from right now. Each year the Government invests about NZD1.4bn in research, of which approximately 40% is spent directly on collecting, creating, analysing and interpreting active research data. Passive data on the other hand, is data that isn't currently being used, or has been collected but not analysed – this is not dead data, simply data that needs to be accessible in case it could be used again in the future. In some cases, "accessible" may mean preservation in a national collection or archive, however in many cases accessibility of passive data simply means it's easy to find out who (which institution, faculty or department) has the data you are interested in.

Access versus Storage

National collections and discipline-specific repositories have a role to play in co-locating data from many institutions, increasing the efficiency of research activities by making data discoverable, and through long-term data preservation. The Royal Society of New Zealand recent report on National Taxonomic Collections suggests there is scope to increase the effectiveness of our national capabilities in preservation and storage of important collections. That said, storage of research data in most cases remains the responsibility of our research institutions. Our research institutions are the employers of our researchers and the holders of research contracts, as such these institutions are also the owners of research outputs such as data, and often have a statutory duty of care when it comes to publicly funded data to ensure that it is value-adding asset for the long term.

It's troubling therefore that many of our research institutions do not feel they have the resources to make data accessible, and in some cases are actively deleting research data due to a lack of data management resource. In 2016, the technology for storing data is inexpensive, but the means and processes for managing data are complex and relatively costly. Without support for data management, our institutions cannot be sure that value created by research funded this year will be accessible for future use. We know from international studies that the expensive, value-adding activities in data management are data acquisition and data sharing.

Quite literally, the "access" and "openness" of research data makes up 85% of the costs of research data management (storage and preservation make up the remaining 15%).

Policy support for research data can also make a big difference to our research institutions; for example, carefully managing and depreciating (relatively inexpensive) storage hardware assets, while fully expensing (relatively costly) data acquisition and sharing each year seems to place the cart before the horse when locating the source of economic value.

The Proposed NRDP

The goal of the National Research Data Programme is to encourage transformative change in New Zealand researcher and research institutions' baseline ability to work with data and employ digital methods. The scale and complexity of this goal requires a coordinated approach that can bring together the complexities of technology, skills, policy and partners in a cohesive, durable fashion.

What the Programme would do, the intermediate objectives, is focus on understanding and implementing the necessary infrastructure and underpinning services that support our research efforts (i.e. NSCs) and enable an effective, impactful research system to achieve our strategic science priorities. We expect the NRDP to be structured as a series of distinct, inter-related, whole-of-sector projects, each focused on a particular data objective, and each with a supporting business case. To make progress, we think a cross-sector, Government-led Programme Implementation Office will be the best approach for scoping NRDP projects and delivering coordinated outcomes for New Zealand.

We think we have a window over the next 5 - 7 years to lift digital research capabilities in New Zealand. Acting across this time period will lower the impact of technology change on the Programme, and ensure capacity and investment in the research sector align with the complementary lift in capabilities and standards across Government and the wider economy.

17

The NRDP is proposed as an intervention with limited goals, and a limited life-span, as means of addressing the scale of the challenge and the need for coordination. This is particularly relevant given current Government focus on New Zealand's data future, the role of data in managing our national assets (e.g. infrastructure, or archives) and the ability of the public services to transform into digital enterprises.

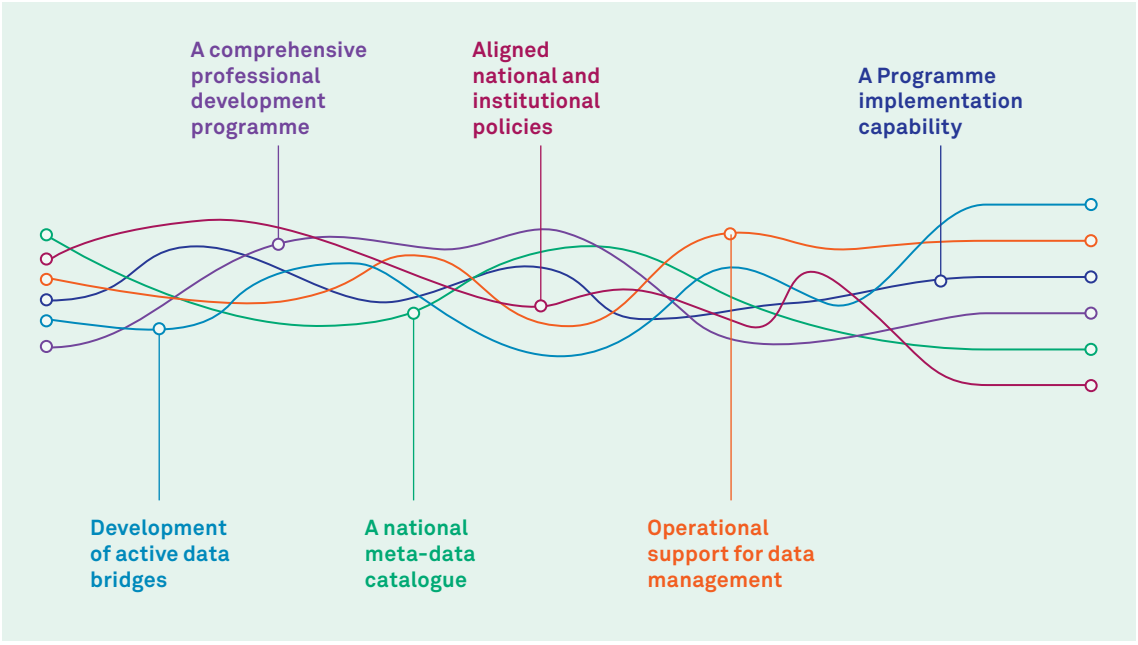
As a programme of activity, the proposed Programme needs to operate against a set of intermediate outcomes that support achievement of the goal.

For the New Zealand research system, we think an implemented National Research Data Programme could:

- Enable higher quality research through a lift in the standards of research, in particular increasing the sophistication and competence in data intensive and digital methods of research, leading to:
  - Better use of the resources and data available to them;
  - Greater collaboration and access to larger scale resources, and
  - Increased participation in internationally funded research.
- Create stronger links between research, government, and industry, to support social and economic development and public sector activities, and in particular contribute to the impact of the National Science Challenges.
- Elevate NZ research sector participation in international / world class research and publishing, and encourage the leaders of our research institutions to engage with global trends in technology driven research.
- Move our research sector beyond funding limited scope projects, towards developing a national capability to sustain a stock of knowledge.

FIGURE 8: PROGRAMME OUTCOMES. THE HIGH LEVEL INITIATIVES WITHIN THE PROPOSED PROGRAMME MIGHT INCLUDE.

Research Data Programme Focus



1__In order to maximise returns on current investment and opportunities, focus intervention on **active research data**, and in particular on the **first use of new data**, as these activities are already supported in the fully-funded science system and will deliver the fastest payback.
Interventions in these areas might include:

   a. Operational support for cross-institutional, research-discipline specific tools and standards for researchers collaborating for data design and meta-data development.

   b. Provision of a national facility for research meta-data to make data discoverable.

   c. Policy changes that incentivise participation in ICT intensive research collaborations, reduce research duplication, and allow institutions to properly value their investments in research data.

2__Develop **active data bridges** between industry, government, and the research sector that allow data to be shared in a trusted, reliable manner. Identify and actively support and extend those that already exist.

3__Provide training and professional development to allow data intensive discovery methods to expand into a wider range of research disciplines.

4__Rework the Nationally Significant Databases, currently cared for by the Crown Research Institutes, into an open, well-managed network of research data that can set the standard for New Zealand research data.

A Multi-Year Programme

Changing researcher behaviour and implementing national programmes are not simple challenges, but we think these goals are worth striving for. Change management is a process that takes time, and bedding in professional development and new skills requires persistence, thus our proposal of a multi-year programme approach. We don't yet know exactly what we need to do, or what each project in a national programme might be, but we believe we need to get started, and we can see momentum building and clarity forming on key early steps.

What happens if we do nothing? If we wait 5 or 10 years, will these changes occur on their own? A quick look back at the last 5 to 10 years suggests not. Unfortunately, in the absence of national goals, the New Zealand research sector tends towards entropy rather than cohesion. The line between competitive tension and collaborative outcomes is seldom clear in research, and this limits our ability to achieve coordinated solutions at economic scale without a national approach. It's fair to say that technology adoption over the last decade has been patchy at best, ignored at worst. We can take heart that we are not alone in this situation – the EU will soon launch a €6bn research infrastructure programme to tackle exactly the same interoperability, fragmentation, and coordination issues we have outlined as challenges for New Zealand. We know that cross-sector leadership can make a difference – our institutions already collaborate on super-computing and genomics, operating centres of research excellence, (and in lobbying Government). This makes it all the more regrettable that the majority of our researchers continue to make do with high-speed network links to their street corner rather than to their fingertips, patchy standards and systems for sharing and managing research data, and few opportunities to scale their research goals or join their data to that of their colleagues.

Our collective experience and observation so far suggests that without a national-level approach to data similar to that seen in other OECD nations, we risk our research system fragmenting further and our research spending delivering less and less value over time. We also weaken the ability of our research institutions to win international research funding, attract international students, and participate in the global digital economy.

Unfortunately, in the absence of national goals, the New Zealand research sector tends towards entropy rather than cohesion. The line between competitive tension and collaborative outcomes is seldom clear in research, and this limits our ability to achieve coordinated solutions at economic scale without a national approach.

2 The Case for Coordinated Action

The Strategic Case for Data

The goal of the National Research Data Programme is to encourage. The strategic opportunity is to transform the quality and impact of research output in New Zealand through a programme of major professional development and training, along with implementation of significant national collections and meta-data curation that enable data co-location and additional data use. This opportunity is available now due the aging demographics of the research sector workforce; the change in cultural attitudes towards data sharing and adoption of digital technologies, and the falling costs of data-related infrastructure and services at a global scale. The way data is considered by international research funders and the heightened expectations of peer reviewers for international publishing are lifting the bar for New Zealand researchers working on the global stage. Beyond this, getting better at valuing and managing research data offers the opportunity to significantly expand our research horizons, do significantly more with the research funding and resources we already have, and intensify knowledge and data transfer between research, industry and the public sector.

"At a national level, New Zealand will not achieve our strategic science goals without a coordinated strategy for research data."

STEPHEN WHITESIDE, CHIEF DIGITAL OFFICER, THE UNIVERSITY OF AUCKLAND

More Impactful Research

In our 2015 Discussion Document, eResearch 2020 observed that there is a significant skills gap in the New Zealand research sector. With the exception of a few research stars who lead their field internationally, our general research population do not have the skills to produce high-quality data-intensive research, nor in many cases to meet the data quality and statistical

standards required in leading international publications. That said, New Zealand researchers as an industry have not had access to the professional development and ongoing training to help them understand these relatively recent expectations. Implementation of a major programme aimed at lifting general research skills, especially focused at those new research leaders who will be directing and spending our public research funding once the baby-boom generation retire, can have a major impact on the quality and currency of New Zealand research and the future of New Zealand. While we know many factors contribute to impactful research, we can be fairly sure that connecting our researchers to comprehensive, well-managed data – data they may not otherwise have access to – will be a key ingredient in lifting the aspirations and scale of the research undertaken in New Zealand.

More Effective Research Methods

Our current research system does not encourage digital efficiency. New Zealand researchers cannot access complementary NZ research or data until after it's published (no sooner than their international colleagues), limiting the scale and impact of every project. Our researchers are geographically scattered, further solidifying barriers to collaboration and sharing, and encouraging duplication of expenditure across (and within) organisations. In their October 2014 briefing to the incoming Minister of Science and Innovation, Universities NZ noted that co-location of researchers produces better research outcomes. Given the rapid advance in data and sharing technologies, national collections and repositories (along with professional data curation and shared data analysis environments) can play a major role in bringing researchers together, as publishing meta-data could help researchers across the country find others doing similar work, or gathering similar data. Such virtual co-location has potential to widen the scope and scale of research goals, reduce the time and cost of research activities, and increase the impact of multiple research efforts and funding streams.

24

The time and cost savings generated through upskilling researchers and providing data curation and sharing through national collections can be reinvested in producing additional research, or in transferring research outcomes to the wider economy or society.

Furthermore, analysis of meta-data in research collections in Europe, Australia, and North America is now allowing for greater targeting of research funding and connecting those researchers and research groups who have complementary interests. Unless we reorganise for digital research and

data management, the efficiencies and research opportunities of "big data" and "cloud" will remain largely out of reach for New Zealand researchers, research institutions or research funders.

Prioritising Standards and Trust

To accelerate our high value industries, we need graduates who arrive in the marketplace already equipped with the data skills they'll need, and familiar with the data standards and expectations of industry - we need our researchers to value these skills and be able to impart them. If collaboration is to thrive across the research sector and into the wider economy, we need our researchers to have similar base skills in data and technology, and to have a basic understanding of each other's tools and standards across research disciplines. Interoperability of data and tools is a huge challenge for researchers around the world; recent New Zealand experience establishing the cross-sector National Science Challenges shows we also struggle to collaborate at the most basic levels of data design and collection.

Like most research sectors around the world, we have our share of research councils, research panels, discipline-specific groups, all of which are effective within their mandate and resourcing – but none of which are mandated towards data. To achieve the scale of transformational change in data capability we are seeking, we need to challenge these groups to provide greater leadership, insight, and technical guidance in data standards (as has recently been the case with the RSNZ report on Taxonomic Collections). In the UK, enabling data and standards in research has allowed the Higher Education and Research Funding Council to propose an entirely new kind of transparency and metrics for research quality. It will not be enough simply to ask for guidance, we also need to empower and resource our research leaders to implement and enforce the standards required. This is important if we're to ensure a high level of transparency in research methods, so that researchers and society can have high trust in research data management and confidence in the outcomes of research.

25

When it comes to standards, the aim is to encourage New Zealand research institutions to participate fully in the emerging norms and expectations for research, and for research management, both for the internal benefit of the research sector and to more ably fulfil the role research plays in the wider society and economy of New Zealand.

Sustaining Research Excellence

International university rankings have become an important scorecard in the global research sector. The rankings are published using data on the quality and quantity of published research, and have a tangible impact on the performance of our research sector in the form of key hires, student admissions, and access to international and commercial research funding. For the foreseeable future, competing universities and nations will have greater access to resources and greater scale than we do; therefore as a sector, New Zealand needs a change that makes a positive impression on our international rankings without beggaring ourselves. It's no longer enough for a faculty or school to have a few research stars publishing in top journals – we will need the majority of our research population (i.e. >50%) to be able to publish at this level if we are to position New Zealand institutions in the upper echelons of the research world. A coordinated approach to research data and digital research will be vital if we are to attract the talent, admit the students, or win the research contracts needed to sustain research excellence.

Demographic Change

In their analysis of census data from 2006, consultancy BERL Economics noted that 62% of all university researchers were aged between 40 and 59 years old, and that the largest cohort of research staff by age bracket was 50 – 54 year olds. Ten years later, that cohort are heading towards retirement and the research workforce as a whole has aged considerably. Researchers don't retire quickly – we're not going to lose experience en masse – however we do have an opportunity to address attitudes, experience and skills of the new generation of research leaders who will be taking up key positions over the next 5 years. Our goals should be to lift the digital maturity of research leaders across all disciplines, rather than letting those already far down the track get even further ahead. Many of the barriers to adoption of base skills in data-intensive discovery will weaken over the coming few years, offering opportunity to accelerate those areas of the research sector, and the economy, that have been slow to adopt and adapt.

An Increasingly Engaged Sector

An eResearch 2020 review across research disciplines and institutions suggested a wide variation in maturity and adoption of digital methods and ICT intensive research. Just as we have some world class researchers scattered across the research sector, we also have pockets of excellence in research data management and digital skills. At the same time, we can also observe increased awareness that research data and digital research methods are challenges we need to take on.

Research data is a uniting issue – a 2015 eR2020 Research Data Workshop included participants from every research discipline, every research institution, every NSC and CoRE, along with government officials, research ICT experts, and institutional leaders. Our research sector might be late to the party in digital and data analytics, yet the sector has the advantage of a highly educated workforce, increasingly led by a new generation across a relatively small industry, with a very high level of connectivity into industry and major public services. All factors that suggest the bang for every buck spent on data capability in the research sector is potentially much higher than elsewhere. What's more, research into similar large-scale shifts in technology, such as internet adoption in the 1990s, tells us that the uptake of new technology by students at university level can rapidly transfer skills into wider society and lower access barriers for those who wouldn't normally be exposed to new technology.

"There are fundamental infrastructural, training and capacity needs to be addressed; access to expertise will become a core capacity in its own right; and standards will be needed in many disciplines for how data is captured and filed, with clear data governance protocols to be established and likely tailored to particular disciplines."

27

SIR PETER GLUCKMAN KNZM, JUNE 2015

Deep Links to Societal & Economic Development

There are key parts of the New Zealand economy and society that will need to rely heavily on advanced data and digitally intensive methods to operate and to advance – a few of which are summarised in Figure 9 below. In many cases, the research sector already has a major role to play in providing national capacity; however in each case, our researchers and institutions struggle with a lack of skills, tools, standards and access to data. The growing flows of data streaming in from sensors in our infrastructure and environment; our homes and cities; and our productive industries suggest our researchers are going to need significant data management and analytical abilities if they're to keep up with the expectations of society and Government.

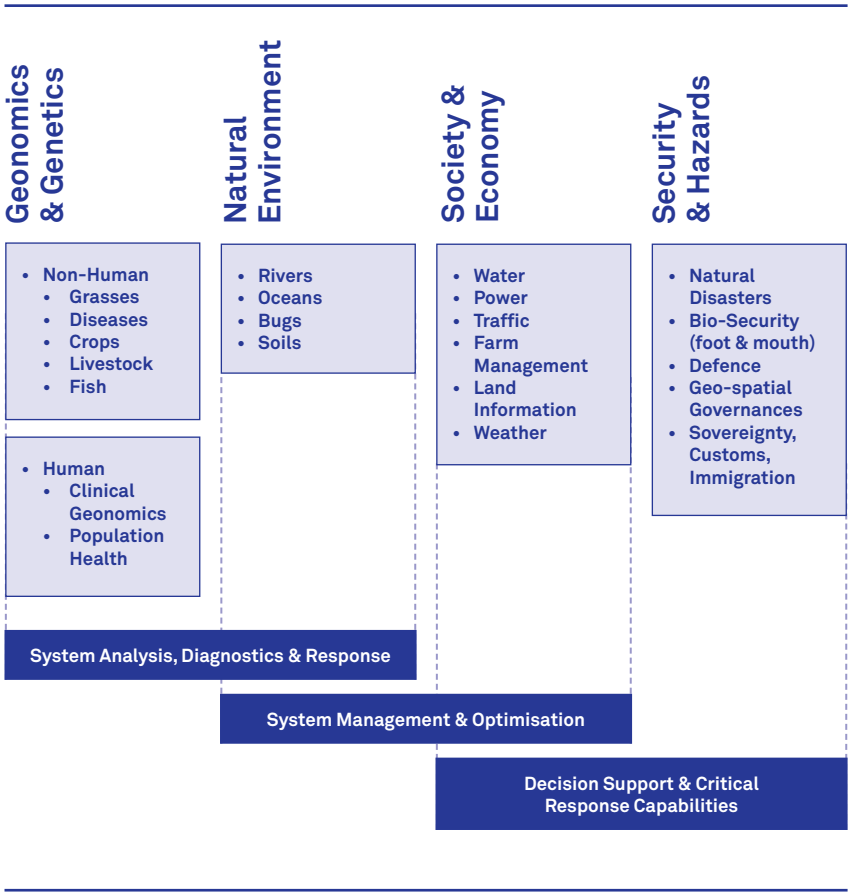


FIGURE 9. ECONOMY & SOCIETAL LINKS TO DIGITAL RESEARCH (ERESEARCH 2020) CHART

A Digital Global Society

The challenges of the future will inevitably have a data and computational aspect to them, be it our response to natural hazards, managing our biosecurity, improving our population health and well-being, or accelerating product and knowledge development in our economy. In February 2015 the NZ Government endorsed the four principles for data that were developed by the New Zealand Data Futures Forum (NZDFF – now the Data Futures Partnership) as guidelines for future development. Each of the four principles has implications for research data and digital research: that NZ should use data to drive economic and social value; that all parts of society should have the opportunity to benefit from data even while individuals retain control over their own information, and that data management should build "trust" in our institutions.

Alongside these foundational values, we think the infrastructure and services required to create, manage, analyse and share research data will underpin many of the opportunities open to New Zealand in the coming decade. We need our research sector to be the trusted experts we can rely on in expressing the values proposed by the New Zealand Data Futures Partnership. eR2020 analysis suggests the value of publicly funded research data could be as much as NZD648m per annum; however our analysis also shows much of that value is unrealised because we're not organised or resourced to capture it.

We don't yet see a research sector fully able to support the national ambitions advanced by Government and the New Zealand Data Futures Partnership – we think we can fix this if we act now.

Economic Case for Data

In 2008, the consultancy Charles Beagrie Ltd published the seminal model for understanding the economic impact of well-managed research data and well-trained researchers. Developed for the Joint Information Systems Committee (JISC) in the United Kingdom (UK), the Beagrie "Keeping Research Data Safe" model has now been applied across multiple European research systems (UK, Ireland, Denmark, The Netherlands, and the EU) as well as in the USA (National Science Foundation), Canada and Australia over the last 5 years. Keeping Research Data Safe includes detailed tools for cost / benefit and value chain analysis across the life cycle of research data. In 2014, Australian economists John Houghton and Nicholas Gruen used the Keeping Research Data Safe model to show that Australian research data is worth between AUD1.9bn and AUD6.0bn per annum to the Australian economy at current levels of research expenditure.

Rather than replicate this work, this proposal takes the initial step of mapping Houghton & Gruen's analysis to the New Zealand context, and then drawing parallels for the New Zealand research system. This means that the economic estimates we include are only indicative – without a detailed NZ study it is difficult to be exact about how much research data is curated and shared, or quantify the inefficiencies that stem from lack of data skills or lack of access to data in research. A next step would be to commission detailed analysis of the New Zealand case; however we consider that step would form an early task for the proposed NRDP.

In Australia, Houghton & Gruen identified two major sources of value from research data:

- **the current value** of research data, generated each year through existing public research investments on a per annum basis, and

- **the potential value** of research data management, of metadata catalogues, standards and tools, collections, curation and sharing that could be realised from data related investments focused on sustaining a stock of research knowledge.

In developing these differing sources of value, our analysis defines an upper bound based on total direct government support for research, per annum; and a lower bound, which is defined as the researcher labour-only costs in publicly funded research, per annum. These bounds are based on figures from the National Statement of Science Investment (2015), on sector assumptions drawn from international practice and set out in Table 1, and on economic analysis conducted for us by PricewaterhouseCoopers (PwC) in Wellington.

Parameter	Value
Annual Value of New Zealand government research spending	$1,410 pa
Researcher labour as share of research spending	48%
Researcher labour cost as share of research spending	$676m pa
Return on investment (Avg. intl. analysis)	40%
Return on investment (NZ Treasury 2006 analysis)	17%
Time spent working with data	46%
Share of research with domestic impact (localisation) – standard assumption	66%
Share of research with domestic impact (localisation) – NZ specific	55%
Depreciation rate	10%
Useful life of data	10 years
NPV Discount rate	8%
Average time savings from data repositories	37%
Researchers who could not have obtained data otherwise	52%
Researchers who could not have obtained data otherwise	52%
Current research data management efficiency (High assumption)	20%

TABLE 1: KEY NZ RESEARCH SECTOR ASSUMPTIONS

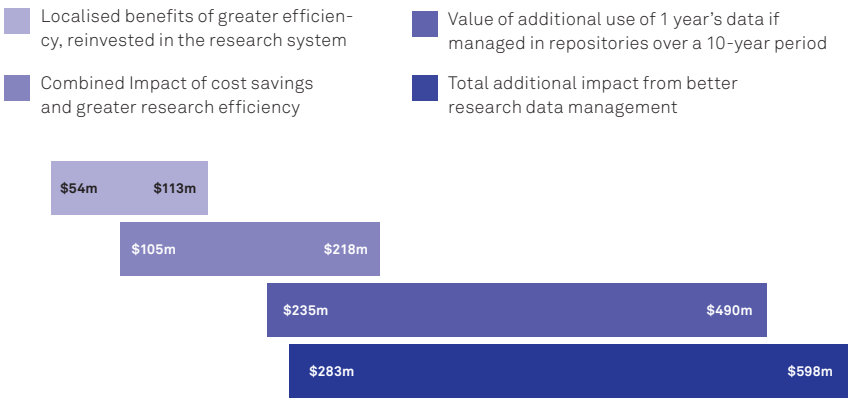


FIGURE 10. THE VALUE OF BETTER RESEARCH DATA MANAGEMENT IN NZ UPPER BOUND FIGURES ONLY

The Current Value of Research Data

Houghton & Gruen present two alternative approaches to valuing data generated through public investment each year; the Use Value approach values data based on the time and cost of producing it, while the ROI on Research Data Activities values the likely return on researcher time.

The Use Value of Research Data

The value of anything when you first purchase it (i.e. an ice cream, a car) is assumed to be equivalent to the cost of the item at the time of purchase. The same can be said of research, and of research data – its value is equivalent to the cost of the activities involved in producing it each year. Analysis in Australia and the U.K. suggests that researchers spend approximately 46% of their time creating, manipulating and analysing data. Applied to the New Zealand research sector, the activity cost or Use Value of research data generated each year in New Zealand through publically funded research is 46% of total expenditure. That's **$283m** pa if we take the lower bound, or **$589m** pa if we base our estimate on all public research expenditure (upper bound).

The ROI on Research Data

$235m **$490m**

The Government invests annually in publicly funded research on the basis that the investment will return benefits to New Zealand society and economy over the long term. Return on Investment (ROI) models differ in parameters for different countries, however the underlying principles are widely understood. As the proportion of total research funding that is spent on creating, collecting, and analysing research data each year is relatively large, we can apply similar ROI calculations to understand the potential returns to New Zealand that might be generated by our annual investment in research data over the long term.

Houghton & Gruen note that in a global research community, only a portion of the value of research activity will actually accrue to the funding country, which are termed localisation returns. If we apply international norms for localisation (66%) and returns on research investment (40%) to the New Zealand context, the ROI on publicly funded research data would be between **$235m** (lower bound) and **$490m** (upper bound) per annum (NB: these figures are lower than the Use Value as we have depreciated the value of research data over only 10 years).

A more conservative approach is to produce a worst case scenario for returns on research data in New Zealand by selecting alternative values for New Zealand. To produce the most conservative estimate of the ROI on research data, we applied a New Zealand Treasury return on research investment rate (17%) which was developed in 2006 to assess R&D productivity in the NZ agricultural sector between 1927 and 2001. We also took into account expectations for the localisation of returns to research effect (55%) in a small country. Even as a worst case scenario, the ROI on 1 year's investment in research data is at least **$56m** pa and possibly as high as **$117m** pa.

33

In developing our ROI analysis, we worked with PwC in Wellington; who noted that public investment in research and research data is not a one-off – we make a similar investment every year, and the returns on those investments will compound over time. Again using the worst case scenario parameters, the present value of a programme of improvement for publicly funded research data is still quite significant at **$110m - $230m**. In calculating these returns, we assumed that improvements would happen gradually over 10 years. If we consider this investment in the optimistic terms of the international model, then compound returns over 10 years of Government investment in research data could be between **$461m - $961m**.

It seems clear that the proportion of annual public research funding that is devoted to collecting, creating and analysing data is considerable, and that the potential returns on this investment over the long term are substantial – even in a worst case scenario. The concern is that, as data-intensive discovery and digital research methods become the norm for global research, our New Zealand research sector find it increasingly difficult to fulfil its potential on the returns to society and productivity that we need.

The Potential Value of Data Related Investments

Houghton & Gruen suggest the potential value of research data related services and the supporting infrastructure can be estimated through assessing the efficiency gains in the research processes which they create (i.e. reduced duplication, use of time, increased collaboration) and the bearing increased access to data has on the effectiveness and impact of research.

Efficiency Gains – Research Funding & Data First Use

Suppose we assume that all researchers in New Zealand have efficient access to data (or at least meta-data) produced by other NZ researchers, or data held by Government or industry, that could be made available for research. What opportunities could be found, what more could be done in such a situation? Efficient access to data or meta-data in Australia and the U.K. has been shown to offer two major advantages to research; the first is that researchers can spend significantly less time seeking (or recreating) data. The second is that research scope and scale that wouldn't otherwise be possible, either due to a lack of resources or a lack of access to data, can now be tackled by a broader cohort of the research community. In economic terms, individual researchers and research institutions become more productive if they have efficient access to data or meta-data. Based on U.K. and Australian experience, efficient access to data can reduce time and cost of research activity by approximately 37%, suggesting a potential productivity gain in the New Zealand research sector of between NZD105m and NZD218m per annum.

Benefits from Additional Data Use – Active Data Bridges & Data Reuse

While this proposal suggests lifting the gains from the first use of new data will have the greatest impact on research, additional use of data has an economic value of its own. Where access to data collections and repositories, along with

researcher skills development, enables researchers to use data they would otherwise be unable to obtain on their own, we assume additional research is produced that could not otherwise have been done (or a similar sized increase in the scale and scope of existing research). Direct extrapolation from Australian calculations suggests the benefits of additional use in New Zealand are likely to be between NZD54m and NZD113m pa.

While these figures are extrapolations from other countries, some with more advanced data-related infrastructure and deeper R&D capital stock, simply adding the efficiency gains and benefits from additional data use suggests that system-wide services such as meta-data cataloguing, data repositories, national collections, researcher skills development and related data infrastructure have **a potential value to New Zealand of between $235m and $490m per annum.**

There is already evidence of good research data management in New Zealand, but we are a long way from our potential. The value of a national approach to research data is that our system could make coordinated progress. If for example we assume that we are (optimistically) 20% efficient and managing our research data on an institution by institution basis today, then a programme such as the NRDP could help lift good data management at each institution over time and deliver additional returns to research productivity of $188m - $392m per annum.

	Labour spending only	**Total spending**
Domestic value of data produced annually in NZ	$283m	$589m
Value from cost savings	$76m	$158m
Value from greater efficiency	$105m	$218m
Value of reusing data	$54m	$113m
Total additional impact	$235m	$490m
Assume 80% benefit	$188m	$392m
10 year total economic impact	$461m	$961m
Worst Case Scenario – 10 yr. total economic impact	$110m	$230m

TABLE 2. SUMMARY OF RESULTS

The Data Investment Required

As early as 2007, the OECD published principles for publicly funded research data for their members, noting in particular that "scientific data infrastructure requires continued and dedicated budgetary planning and appropriate financial support. The use of research data will not be maximised if access, management, and preservation costs are an add-on or after-thought in research projects. It is important to note, however, that the cost of storing and managing data has decreased dramatically in recent years, and lack of knowledge about such changes can, in itself, be a barrier to advancement". While it's fair to say that the costs of storing and managing data have continued to fall since 2007, the barriers to advancement from a lack of knowledge still appear quite firm.

The work by Beagrie on Managing Research Data in the UK was followed by further projects, and development of modelling tools and factsheets that offer some key insights regarding the cost dynamics of research data management, broadly summarised in Table 3 below. These tools for cost/benefit analysis and risk management were further developed by the European Commission "4C Project" which road mapped options and analysis tools for sustainable data management across the EU by 2020. As with the ROI analysis in the previous section, we do not seek to replicate these studies here, but instead draw broad parallels that can inform work by the proposed Programme in due course

TABLE 3. KEEPING RESEARCH DATA SAFE FACTSHEET SUMMARY

	Insight	Implication
Activity Costs	Acquiring and ingesting data costs the most. Archival, storage & preservation activities cost very little by comparison. Early acquisition & ingestion of data lowers total costs over the life cycle.	The largest potential cost efficiencies will come from tools and training that enable researchers to manage and log their data easily and early in the research process.
Fixed Costs	Long-term fixed costs do not vary with the size of collections. Fixed costs are generally staff & skills related, and can fall per-unit cost can fall over time.	Economies of scale from multi-institutional collaboration and outsourcing are worth pursuing.
Cost Trends	Cost fall over time. Technology capability increases while technology costs generally fall. High set up costs do not translate into high refreshment costs.	Implies that over time, data archives and management become sustainable activities.

Technology dynamics have changed the potential cost structures of data related infrastructure and services cited in the earlier studies. In estimating costs in Australia, Houghton & Gruen directly applied the UK experience of between 1.4% and 1.5% of total public research expenditure. A direct extrapolation to New Zealand would suggest costs of NZD21m per annum (based on total research expenditure), however we suggest these figures are likely to be misleading for three reasons:

Small Country Premium

New Zealand investments seldom achieve economies of scale or geographic co-location of equivalent overseas investments. Typically, we pay a premium on like-for-like investments in order to optimally serve a small, geographically fragmented research population and therefore best realise the potential of the investment. While comparative data would be needed to assess this premium, it might be as high as 20% during initial programme development, increasing annual costs by NZD4.2m.

Catch Up Costs vs. Late Adopter Advantages

New Zealand researchers are starting behind the pack in terms of skills and access, therefore costs are likely to be higher in the near term when there will be increased focus on training and development across the sector. That said, New Zealand may benefit from late adopter advantage in many of these areas. Training and professional standards in many areas of data-intensive research have already been developed overseas. International standards and resources for skills and access have become well-established over the last 5 years, and associating ourselves with these existing resources (i.e. ORCID) should be considerably less expensive now they are up and running. Overall, we think the late adopter advantages have the potential to fully offset any catch up costs, provided the NRDP is well managed and our research institutions are aligned on their goals for research data in New Zealand.

Falling Technology Costs

The pace of technological change in cloud services and data technology is driving costs down and lifting capacity at an astonishing rate. At the same time, the number of researchers working with data and the quantity of data generated is also on the rise. These factors will not only affect operational costs in terms of people and training, but also improve access to economies of scale in data technologies for the New Zealand research sector as a whole – provided coordinated sector investments can be made. We think well managed, timely investment should be able to take advantage of compound annual savings of at least 10% of total costs over the first 5 years of the programme, suggesting total savings of up to NZD10.5m through falling technology costs.

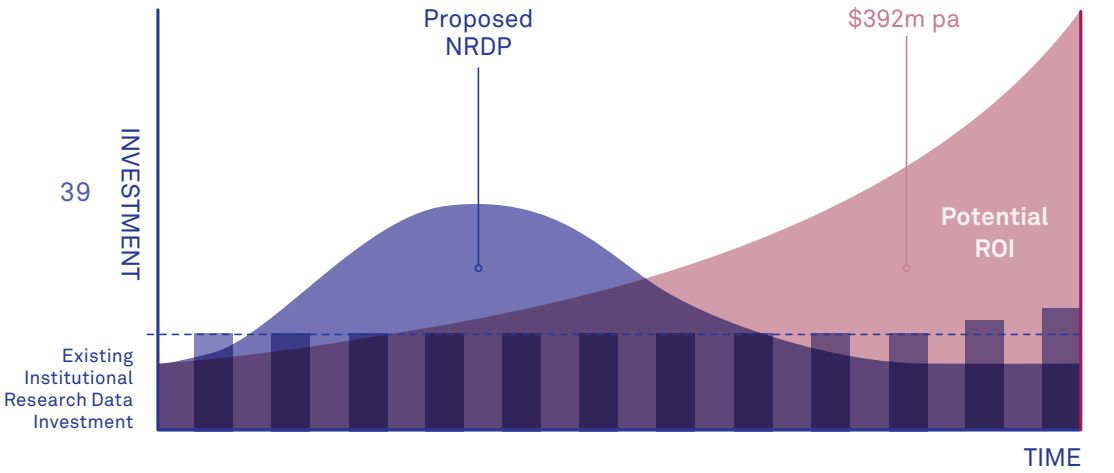
Funding Options for Coordinated Action on Research Data

Taken together, these cost figures suggest annual investment across the NZ research system of approximately NZD15.1m per annum would support research data related services and infrastructure, and lift researcher digital skills development in New Zealand. Considering that our total annual investment in creating, collecting and analysing research data is about $300m, costs of $15.1m per annum to sustain that stock of knowledge don't seem unreasonable. Unfortunately, eResearch 2020 estimates New Zealand research institutions currently invest $6m - $8m per annum in research data management activities (mostly data storage, curation, and digitisation), or 0.5% of total public sector research expenditure. This would suggest a funding gap in 2016 of $7.1m to $9.1m; however, we estimate this institutional spend will grow by up to 30% by 2020 as rapid growth in data generation and digital research stretch the resources of our institutions. Given that

total public research expenditure is not expected to grow at this rate, the gap between New Zealand investment levels and international benchmark investment levels will narrow. Unfortunately, our existing investment is uneven across institutions, is focused on narrow institutional priorities, and is fragmented and duplicative across research disciplines. Given the substantial benefits of good research data management, and the volume of under-realised potential in our research system, we think we have a real opportunity to get significantly more value from this annual 0.5% of public research spending.

The NRDP has been designed as an opportunity for Government to provide short-term for uplift in research data management across the New Zealand research sector. As such, the NRDP would lift total research data management funding to 1.5% of public research expenditure on average over the course of 5 years, with Government making up funding the difference over this time period. Based on the investment gap and the potential returns, we propose Government allocate $45m - $50m in total to fund a NRDP over a 5-year period aimed at super-charging the existing investment our research institutions are already making in research data. The Government's NRDP investment would be towards design and implementation of 5 major workstreams, and to develop key, missing pieces of the research data eco-system.

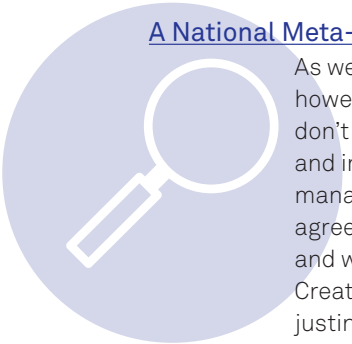
FIGURE 11. NET PROFITS (COST/TIME/ROI)



3 NRDP Initiatives

Major Programme Outcomes

The big building blocks of a programme of change have been identified by the research community through workshops and interviews over 2014 and 2015 and drive the intermediate outcomes we think will lead to achieving the true objective of transformational change across the entire system. These major aspects will need to be tailored so they're complementary to aligned Government initiatives in non-research areas, and refined to fit the exact needs of the New Zealand system; however, in each case we seek to prioritise impact from the first use of new data.



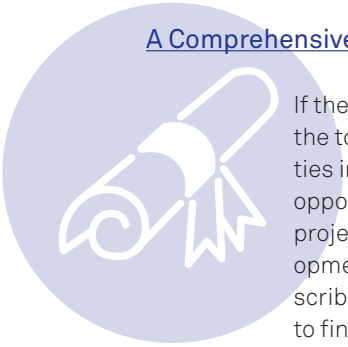
A National Meta-Data Catalogue for Research Data

As we've noted, data doesn't need to be free to be open and accessible, however data does need to be discoverable. Right now, our researchers don't design their data to be findable, let alone searchable. Our institutions, and indeed our research funders, do not have a consistent approach to data management or sharing meta-data. Our research communities do not have agreed meta-data standards for creating, collecting or sharing new data, and we often lack the tools to fully participate in international collaborations. Creating a meta-data catalogue across all New Zealand institutions, and adjusting incentives to help researchers manage meta-data, could realise much of the untapped value in our research system.

A national home for research meta-data is not the same as a repository, storage would remain the responsibility of research institutions (while meta-data technical standards need to be governed by research discipline communities). Such a catalogue would allow traceability and measurement of data value created, and of additional use of passive data. New Zealand currently lacks a deputised cross-sector agent that can engage with international research data efforts on behalf of NZ research, an implemented national research data facility would correct this. Even greater potential exists for a national meta-data catalogue to underpin new or ad hoc research initiatives such as NSCs or regional institutes, reducing duplication and contributing to the effectiveness of research investments. As noted earlier, coordinated strategy for research data really needs to be the starting point for research collaboration and multi-institution research efforts.

Where research is conducted across institutions and alongside industry, such as with the National Science Challenges, a national catalogue facility might also provide a unified home and structure for the data that emerges from the collaboration over the decade-long research journey, without developing dedicated structures around each endeavour. Examples of this approach can be seen in Europe, and most recently in the United States with the establishment of the US$500m Materials Genome Initiative – a discipline-specific response data infrastructure funded by the federal Government to support a major scientific challenge in materials science, but hosted and operated by a cross-sector consortium – the US National Data Service. Once such a national mete-data catalogue is established for New Zealand, providing ongoing support to major research efforts could be an important factor modelling the sustainability of the facility. We think our university libraries and CRI data programmes will be well placed to take on ongoing management and development of this facility. CONZUL, the Council of New Zealand University Librarians, highlighted in their recent Universities NZ report, "Managing Research Data," both the need for a national approach to meta-data and data discoverability, and the willingness of the research sector to progress this goal. CONZUL are already making efforts to understand the feasibility of such a national catalogue and how researchers, institutions and research publishers would interact.

A Comprehensive Professional Development Programme



If the time is ripe for the research sector to leap forward, we need to provide the tools and opportunities for our researchers to excel. The research activities involved in the first use of newly collected or created data is an important opportunity, as it suggests the possibility of getting additional value from projects or research streams that are already funded. Professional development needs to improve the way we design our research, structure and describe our research data, and shape our research goals. We should know how to find other researchers who are also funded to do complementary research to our own, and we should know how to access major infrastructure that can expand the boundaries of our research scope.

While there's no doubt the research institutions will need to play a big role in any professional development of their staff, we can expect that accelerating change across the entire industry to the next level of capability will require a cohesive, national approach. New Zealand researchers are not the only ones who face the challenge of learning new skills for new, digital research methods. The European Commission will soon announce a €6bn European Open Science Cloud initiative, much of which is aimed at changing researcher behaviour, increasing data interoperability, and overcoming researcher fragmentation. Much of the development we're suggesting is required can

be found in online training modules, nano-degrees, and digital discovery, the role of a national programme would be to engage researchers, coordinate access and standards, and to agree assessment protocols and goals. Key Programme partners in developing and implementing metrics and expectations at a national level might be the pan-sector research governance committees at Universities NZ and Science NZ. A big push in sector-wide development does not go on forever – once the next tier of capability is reached by an early majority and a degree of momentum is established, conventional Continuous Professional Development Plans can take up the ongoing responsibilities of training.

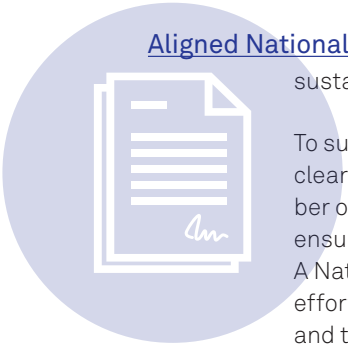
Operational Support for Data Management

Research discipline specific communities across institutions need to be the ones that govern the technical standards for creating research data (in conjunction with international research and industry standards where these exist). To date, these communities and technical standards have not appeared organically. Our researchers and institutions need help to come together, agree and implement the necessary technical standards and tools that will drive meta-data creation and sharing.

The goal is that shared tools and standards for data from the outset of the research process, both within research and across disciplines and industries, will improve interoperability of research data from different sources, easing collaboration and ultimately more impactful research outcomes. A strong understanding of the expectations of industry will lead to graduates who are better equipped when they emerge from study. We already have a number of permanently instituted, research-discipline specific committees, mostly focused on making funding decisions; however, we need to permanently establish Technical Working Groups between our universities and CRIs that guide meta-data standards, data analysis and tools both within the research discipline and the interface to industry. While this is an ongoing operational element of the programme, much of the institutional policy foundation for setting up these technical working groups has been laid by the Royal Society of New Zealand, and research sector initiatives such as the Lincoln Hub Data2 group. Resourcing our existing groups to provide an evolving and ongoing level of guidance across the sector will have immediate impact on cross-sector collaboration such as CoREs or NSCs, and ensure the other Programme initiatives have a touchstone for engagement and development across the broadest base possible.

Development of Active Data Bridges

Social license, privacy, confidentiality, and ownership issues all play an appropriate part in developing the value of data. As research and public-sector data become increasingly linked and aggregated, the research sector needs to ensure that society can have confidence in the ability of researchers and research institutions to adequately manage their research data – not only for privacy, but also for research value. New Zealanders should expect greater transparency from research institutions on their research data management practices, and measurement of how effective these practices are at generating benefit to New Zealand and at safeguarding confidential data. The NZ Data Futures Partnership (formerly the Data Futures Forum) is developing a framework for New Zealand data in society, and the research sector has a significant role to play by creating structured, trusted access for data to safely change hands. Trusted data bridges already exist to some extent, for example between Landcare Research and the Ministry for Primary Industries in biosecurity, or between AgResearch and Livestock Improvement Corporation in herd genomics; however, these are one-way bridges, and only scratch the surface of what could be achieved. A meta-data catalogue and a higher general skill level are starting points for expanding data capability into society and economy, and our existing PGP and NSC programmes are good candidates for new, permanent data bridges. A data bridge could as simple as a set of agreed protocols for safe data sharing, or as advanced as dedicated, shared storage and curation facilities – more of either is desirable. Callaghan Innovation has already extended an invitation to help prototype NRDP initiatives, and developing a proof of concept "active data bridge" between a research community and an industry segment could be a tangible first step. For a national investment in data capability to really pay off, we need to ensure that the capability flows into society, and has a measurable,



Aligned National & Institutional Policies

sustainable impact on the wealth and welfare of New Zealanders.

To succeed as a programme of change, our researchers and institutions need clear incentives and expectations. In the transition to digital, there are a number of policy areas to coordinate, not least the engagement with public and ensuring the sector has the social license to exploit data for knowledge. A National Research Data Programme would complement other national efforts in data – including Statistics NZ (e.g. the NZ Data Futures Partnership) and the Department of Internal Affairs (e.g. the Digital Preservation as a Service initiative), and better enable the sector to meet statutory expectations such as the Public Records Act 2005.

Specifically in the research sector clear expectations and incentives probably include a common legal framework for data across institutions, and an appropriate accounting treatment for data assets in research institutions to ensure data retains value beyond the financial year it was created in. There are also policy levers that can begin to influence researcher behaviour – a common policy for research data management across all contestable research funding would be helpful, as would a requirement for researchers to specify and share indicative meta-data for the agreed research during the contract negotiations phase of funding.

Finally, a centralised approach to accessing enabling technologies and licenses for tools might not only access economies of scale but might also reduce platform divergence and investment duplication across the sector. Some of these might be as transactional as joining New Zealand to the International Foundation for the Digital Object Identifier (DOI) standard that identifies data resources for citation, or adopting the business model and meta-data guidelines provided by international organisations such as Datacite. At the other end of the scale, the NRDP might champion change through rolling out the unique researcher-focused ORCID digital identifier at a national level. ORCID is a not-for-profit international standard that aims to increase trust and transparency in data and research, potentially delivering new methods and metrics for governance, peer review, and collaboration in research. Much of this could be seen as policy tinkering and procurement alignment, yet clearly stated goals and incentives to change can help the sector to progress rapidly and evenly. The role of the programme will be to assess each of these initiatives at a national level, and to execute on those that the sector agrees can offer the best outcomes for New Zealand.

A Programme Implementation Capability

As with any long-term, transformational opportunity, we suggest the first step would be to put in place Programme Management that can shape and implement the various projects to be undertaken as part of programme delivery. This Programme Management function will need to be an effective change agent in both the public service as well as across the research sector, and have the skills to develop and deliver complex projects. Depending on structure and the complexities of the projects to be undertaken, programme management could be expected to require between NZD4.0m and NZD6.0m over the life of the programme.

In general, we would expect a National Research Data Programme to start small, with initial projects that focus on policy and alignment opportunities to begin with, perhaps along with engagement with those readily accessible international tools such as ORCID and the Datacite DOI standard. As this is a long term programme of change, individual projects will need to be shaped into digestible investment chunks, with the appropriate business case and returns on investment defined at each stage, and then risk managed through implementation. We suggest a reasonably standard distribution of total investment across the entire 5 year programme, with the heftiest projects worked up and delivered in the middle years, and the programme winding down over the final 18 months.

NRDP Structure

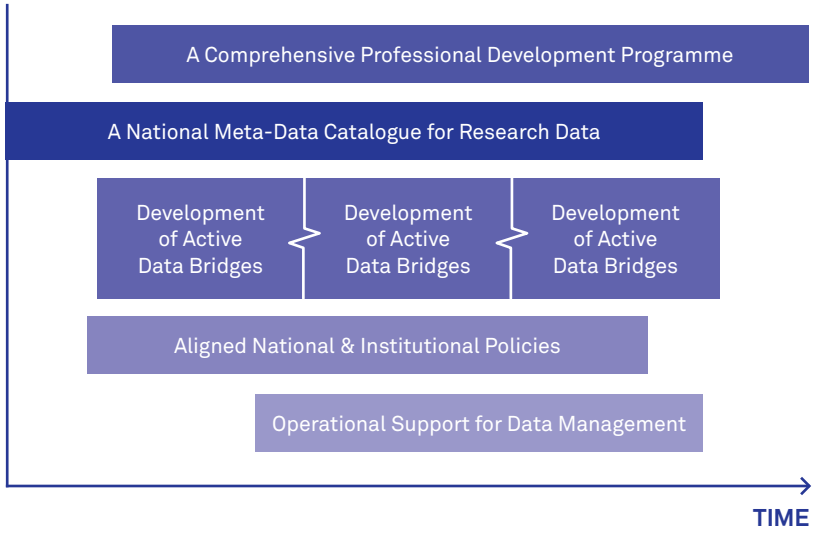


FIGURE 12. NRDP STRUCTURE

Engaged with the Community

A key aspect of a successful National Research Data Programme will be the extent to which it is tied to existing leaders in the research community, and develops the New Zealand research data eco-system. In developing the NRDP workstreams, we expect natural partners will emerge for each initiative, much in the same way CONZUL have taken up the baton in breaking-ground on the feasibility of a New Zealand-wide research meta-data catalogue. Likewise, Crown Research Institutes and Callaghan Innovation might become major actors in establishing and operating Active Data Bridges between research, industry and Government. We propose the NRDP as an opportunity to establish some of the missing pieces of our NZ research data eco-system, and to help key actors in NZ research and industry to excel.

eResearch 2020 Recommendations

We've designed the National Research Data Programme based on input from researchers, research institutions, and through engagement with public services and the international research data management community. While we've proposed the NRDP in some detail, what we are suggesting is a broad programme of change. Individual projects will still need to be shaped and implemented, therefore the first step will be for Government and the sector to get behind establishment of a Programme Implementation capability. We recommend the following:

For Government

Establish and fund the Programme Implementation workstream, either within the public service or in partnership with the research community, and bring the NRDP to life across the New Zealand research sector.

For Research Institutions

Begin reorganising for digital research and data-intensive science, and ensure senior management participation in the development and implementation of the National Research Data Programme nationwide.

For Researchers

Broaden your engagement with digital methods, and increase your demand for research data participation from your colleagues, your research and your institutions.

PROFESSOR PETER HUNTER, ER2020 2015

"Inevitably, our society's problems in the future are going to have a data and computation aspect to them."

References

- BERL Economics: ***Academic Workforce Planning – Towards 2020,*** Report to the Universities NZ HR Committee Steering Group, November 2010, http://universitiesnz.ac.nz/files/Academic_Workforce_Planning_Towards_2020_FINAL.pdf
- Catlin et al, ***Raising your Digital Quotient,*** McKinsey Quarterly June 2015 http://www.mckinsey.com/insights/strategy/raising_your_digital_quotient
- Beagrie N., Chruszcz J. & Lavioe B., Keeping Research Data Safe, JISC 2008 http://www.beagrie.com/krds.php
- Charles Beagrie Ltd & JISC, Keeping Research Data Safe Factsheet 2010, http://www.beagrie.com/krds.php
- CONZUL Working Group, R***esearch Data Management Framework Report,*** Universities New Zealand, November 2015, http://www.universitiesnz.ac.nz/node/855
- Datacite – International Data Citation Initiative, ***Business Model Principles,*** Datacite 2012, www.datacite.org
- eResearch 2020 Discussion Document, eResearch Challenges in New Zealand, 2015, www.eResearch2020.org.nz
- ***European Commission, European Cloud Initiative – Building a competitive data and knowledge economy in Europe,*** Brussels, draft 2015, http://www.politico.eu/wp-content/uploads/2016/03/SKM_C454e160303162102.pdf
- Goldfarb A., ***The (Teaching) Role of Universities in the Diffusion of the Internet,*** University of Toronto, 2004 http://ftp.zew.de/pub/zew-docs/div/IKT04/Paper_Goldfarb.pdf
- Higher Education Funding Council for England, ***The Metric Tide: independent review of the role of metrics in research assessment and management,*** July 2015, www.hefce.ac.uk/rsrch/metrics
- Houghton J & Gruen N, ***Open Research Data: A report to the Australian National Data Service 2014,*** www.ands.org.au
- International DOI Foundation, ISO 26324, www.doi.org
- Kaye-blake B. PwC eResearch 2020 letter final, PwC, Wellington, March 2016, http://www.eresearch2020.org.nz/wp-content/uploads/2016/03/PwC-eResearch-2020-letter-final.pdf
- MBIE, ***Draft National Statement of Science Investment 2014 - 2024,*** May 2014, www.mbie.govt.nz
- MBIE: The National Statement of Science Investment, 2015 - 2025, August 2015, www.mbie.govt.nz

- MBIE, ***Regulating communications for the future: Review of the Telecommunications Act 2001,*** September 2015 www.mbie.govt.nz
- McKinsey Global Institute, ***Digital Globalisation: the new era of global flows,*** March 2016, http://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/digital-globalization-the-new-era-of-global-flows
- National Infrastructure Unit, ***The Thirty Year New Zealand Infrastructure Plan 2015,*** http://www.infrastructure.govt.nz/plan/2015/nip-aug15.pdf
- New Zealand Data Futures Forum, ***Key Recommendations and Catalyst Projects,*** February 2015, https://www.nzdatafutures.org.nz/sites/default/files/NZDFF_Key_recommendations.pdf
- New Zealand Data Futures Forum, ***Harnessing the economic and social power of data,*** February 2015, https://www.nzdatafutures.org.nz/sites/default/files/NZDFF_harness-the-power.pdf
- NZIER, ***The Economic Value of University Investment,*** October 2010, www.universitiesnz.org.nz
- Office of the Press Secretary, ***President Obama Launches Advanced Manufacturing Partnership,*** The Whitehouse, June 2011, https://www.whitehouse.gov/the-press-office/2011/06/24/president-obama-launches-advanced-manufacturing-partnership
- ORCID (Open Researcher and Contributor ID), www.orcid.org
- Organisation for Economic Co-operation and Development, ***OECD Principles and Guidelines for Access to Research Data from Public Funding,*** OECD 2007, http://www.oecd.org/sti/sci-tech/38500813.pdf
- PMCSA, ***"The intended and unintended consequences of e-research: why scientists must engage openly with the community".*** Keynote address by Sir Peter Gluckman KNZM, eResearch 2020 Workshop, Wellington 2015, http://www.pmcsa.org.nz/wp-content/uploads/Speech_e-research-2020.pdf
- Royal Society of New Zealand, ***National Taxonomic Collection in New Zealand,*** December 2015, http://www.royalsociety.org.nz/expert-advice/papers/yr2015/national-taxonomic-collections-in-new-zealand/
- Scobie G.M. & Hall J., ***The Role of R&D in Productivity Growth: the case of agriculture in NZ 1927 – 2001,*** NZ Treasury Working Paper 06/01, March 2006, http://www.treasury.govt.nz/publications/research-policy/wp/2006/06-01/twp06-01.pdf
- The 4C Project, ***Investing in Curation: a shared path to sustainability,*** http://4cproject.eu
- The Materials Data Facility Website https://materialsdatafacility.org/
- Universities NZ Briefing to the Incoming Minister, October 2014, http://www.universitiesnz.ac.nz/node/765

Design: Studio Alexander

eResearch 2020

Visit our new website: **eresearch2020.org.nz**

If we consider our socio-economic well-being to be linked in part to our research sector and our ability to understand the world around us, then arguably a lag in eResearch adoption may create limits to our productivity, our social cohesiveness, or to our capacity to monitor our environment, our borders, or our economy.