MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

Research, Science and Innovation Data: Conceptual Model

Summary of Submissions

14 June 2017

New Zealand Government

List of Abbreviations and Acronyms

CoRE	Centre of Research Excellence
CRI	Crown Research Institute
HRC	Health Research Council of New Zealand
IRANZ	Independent Research Association of New Zealand
MBIE	Ministry of Business, Innovation and Employment
MfE	Ministry for the Environment
MPI	Ministry for Primary Industries
NRIS	National Research Information System
NSC	National Science Challenge
NSSI	National Statement of Science Investment
NZAS	New Zealand Association of Scientists
NZBN	New Zealand Business Number
OPMCSA	Office of the Prime Minister's Chief Science Advisor
ORCID	Open Researcher and Contributor Identifier
RS&I	Research, Science and Innovation

Contents

What is the Research Science and Innovation Data Conceptual Model?

The Research, Science and Innovation (RS&I) Data Conceptual Model is an important part of developing the National Research Information System (NRIS)[1] for New Zealand. The Conceptual Model helps define the scope of the data that the NRIS could eventually contain and starts the process of defining a set of common data standards including definitions for concepts and data elements, accompanied by guidance for use.

Creating the NRIS is a key action in the 2016 Research, Science and Innovation Domain Plan[2]. The Domain Plan provides the strategic direction and key actions required to support the vision of the National Statement on Science Investment, specifically "easily available, reliable data on the science system"[3].

The Domain Plan also calls for common data standards across funding agencies and the research community. Common data standards will help improve data quality, reusability and interoperability across New Zealand's RS&I system and facilitate the creation of the NRIS.

The NRIS will enable aggregation of data and comparability across different parts of the system, increase efficiencies around reporting, and reduce transaction costs. By communicating the scope of data encompassed by the NRIS, and the associated data standards, the Conceptual Model helps everyone plan information systems and envision possible uses for the system.

MBIE led the development of a draft Conceptual Model, working closely with RS&I agencies and key stakeholders. This included targeted consultation and workshops with funding agencies and the research community. The draft model was released for public consultation.

[1] For more information on the NRIS project see http://www.mbie.govt.nz/info-services/science-innovation/research-and-data/nris
[2] For more information on the 2016 Research, Science and Innovation Domain Plan see http://www.mbie.govt.nz/info-services/science-innovation/research-and-data/sector-data
[3] For more information on the National Statement on Science Investment see http://www.mbie.govt.nz/info-services/science-innovation/national-statement-science-investment

What is the purpose of this Summary

A consultation document[4] on the draft Conceptual Model was released on 3 March 2017 and **submissions closed on 14 April 2017**. For ease of reference, the Level 1 and Level 2 diagrams from that consultation document are repeated in Appendix 2 of this document.

This summary document reflects the feedback MBIE took from the submissions. The document lists who submitted and sets out the high level themes emerging from the submissions. A brief summary of each submission and a collation of comments on each entity are included as appendices.

The inclusion of a submitter's comment in this document is intended to help illustrate points made by submitters. No inference as to MBIE's views on the issues raised ought to be taken from the inclusion or exclusion of a particular comment or opinion.

MBIE will use this feedback, and our ongoing engagement with the sector, to help develop the first approved version of the Conceptual Model.

What are the next steps?

What is MBIE's planned approach?

Communication and engagement are critical for developing the National Research Information System (NRIS). Some of the issues raised in submissions are about the nature of the NRIS project as a whole rather than just the Conceptual Model. MBIE will be engaging with stakeholders about the collaborative rollout of NRIS over the next few years. Implementation work will take several years and MBIE will maintain an ongoing conversation with the sector about the NRIS.

As part of this, MBIE will continue to engage with the RS&I Sector on the Conceptual Model. The first approved version is likely to be available in late August, and discussions with submitters and other stakeholders are ongoing.

[4] The consultation document can be found here: http://www.mbie.govt.nz/info-services/science-innovation/research-and-data/pdf-library/rsi-data-conceptual-model-consultation-draft-march-2017.pdf

The Conceptual Model is a guide to the scope of data that could be included in the NRIS when it is fully realised. MBIE will work with the RS&I community to gradually bring data from institutions into NRIS. At present MBIE's focus is on working with funders to bring their data on board and identify any issues with the model as this process unfolds.

The Conceptual Model is part of the ongoing NRIS conversation. A key role for the Conceptual Model is to help institutions plan for future data sharing so that this is easier when the time comes. Equally, feedback from the sector is critical in ensuring we evolve a model that is flexible enough to encompass diverse funding systems and the range of pathways research can follow. The first approved version of the Conceptual Model will focus on aspects that we think are stable and can assist others with planning as well as guidance on data elements that are necessary to enable effective data integration.

Two areas go to the heart of creating broad acceptability for the NRIS: identifying value the whole sector can derive from the NRIS and ensuring sound governance (including confidentiality of data). MBIE is working with stakeholders to understand and articulate these issues, and generate appropriate solutions. MBIE will publish material on these at the same time as the first version of the Conceptual Model. This will help frame the NRIS project as well as making clear the role of the Conceptual Model within the wider NRIS framework.

What is the proposed timetable for further work?

June to mid-July 2017	Ongoing engagement and discussion around issues raised in submissions, evaluation of options to refine the Conceptual Model, development of supporting documentation
June 2017 onwards	Individual and group meetings to discuss NRIS development pathway
July-August 2017	Finalise and approve next version of the Conceptual Model and supporting documentation
Late August 2017	Release of next version of the Conceptual Model and supporting documentation

Who submitted?

Submissions were received from the following groups, organisations and individuals:

Business-related Industry Bodies and Private Sector	**Tertiary Education**
Business NZ	The University of Auckland
Digital Science	Auckland University of Technology
	University of Waikato
CRIs	Massey University
Science NZ	Victoria University of Wellington
Landcare Research	Lincoln University
Plant and Food Research	University of Canterbury
	University of Canterbury Library
Peak Science and Funding Bodies	University of Otago
NZ Association of Scientists	
Callaghan Innovation	**Central Government Agencies**
Health Research Council (HRC)	Ministry for the Environment
KiwiNet	Ministry for Primary Industries
Marsden Fund Council	Office of the Prime Minister's Chief Science Advisor
Royal Society of New Zealand	Statistics NZ
Other	
Herve Thevenon	

The Independent Research Association of New Zealand (IRANZ) provided verbal advice that they were happy with their participation in the process and had no comments on the draft model. Other stakeholders who may be interested in future developments but did not submit include polytechnics and institutes of technology, private firms, industry sector organisations, Māori research organisations and individual researchers.

What was the overall response?

Submissions generally canvassed a wider range of topics than were covered in the consultation document. Many submitters took the opportunity to provide comments on various aspects of the overall NRIS project, as well as the specifics of the Conceptual Model set out in the consultation document.

Submitters were generally supportive of the idea of an NRIS and recognised the draft Conceptual Model as a solid start, while also posing questions about specific aspects of the model and the relationships within it. Some submissions reflected uncertainty around implementation, including links to the broader NRIS project and other initiatives such as ORCID. Many discussed potential impacts on their organisation. Some submissions were brief and focused; others were wide ranging.

About 50% of submitters used the questions in the discussion document to structure responses. The others commented using their own structure. Most submitters provided general comments. All but two provided feedback on the core concepts in Section 1 of the Consultation Document, and most provided feedback on the elements and entities (i.e. definitions and data requirements) in Section 2. Just over half of the submitters provided comment (some very briefly) on the code sets in Section 3.

Around half of the submissions also suggested types of worked examples that could be used to illustrate how different aspects of the RS&I system are captured by the model. There is a moderate degree of overlap in these suggestions.

Areas of uncertainty for submitters in relation to the Conceptual Model, and the NRIS generally, are summarised in the next section under four broad themes:

- Compliance
- Comprehensiveness
- Comparability and Compatibility
- Confidentiality

What are the high level themes across submissions?

Support

Most submitters were supportive of the direction and scope of the Conceptual Model.

> *"The model, and the implied National Research Information System, represent the potential for significant improvements in the ability to view, understand, and assess the performance of the New Zealand research system."*

They identified the potential for *"…simplifying reporting across the sector…"* and *"…providing a more rigorous evidence base to document the source attribution and impact of New Zealand's research base."* "*The purpose and intent… is laudable"* noted one submitter, while another stated *"….we support the collection of good information that will in turn give greater clarity and visibility of where NZ gets the best return on its R&D investment…."*.

Submitters acknowledged that the model was a solid initial attempt to capture the Research, Science and Innovation system within New Zealand. One observed that *"The conceptual model laid out in the document does seem to represent the system as it operates at a high level and it is good to see the model driving consistency in descriptions and definitions across the system"* and another described it as *"...pretty comprehensive"*.

Comments about engagement and consultation so far were positive; one submitter welcomed *"the consultative approach that led to the creation of the conceptual model and the involvement of .. staff throughout the process"* while another expressed their willingness *"to be involved as you further develop the proposal, and …. [to continue] to work with you on this initiative."*

Uncertainty

Some submitters are **uncertain** about both the way the draft model represents aspects of the RS&I system and its fit with their processes at a detailed level as well as the implications of the NRIS for their organisation. Comments often addressed both the specifics of the model included in the consultation document and wider issues about NRIS implementation. For submitters, the model and its implementation in the NRIS were closely related; this gave rise to questions about how the NRIS will be implemented as well as what it may include and why.

Several submitters suggested a phased roll-out of the NRIS, working with those who will provide data and involving *"…pilot data collection exercises across different research providers and RS&I agencies"*. One submitter commented that they have *"…few issues with the data you are planning to collect and see… it as a valuable resource. However the timing may be an issue…."*.

Submitters who were concerned about the way the model represents the RS&I system and the scale of data required were also *"…supportive of a staged roll-out of the RS&I data model…over a number of years"* as they would be able to *"supply the data far more readily once [a] new research management system comes online"*.

Many submitters wanted greater clarity around timing of actions and the role of, and impact on, submitting institutions. Some submitters provided high-level comments about aspects of the model they felt could be improved as well as detailed feedback on the entities, data elements and relationships in the draft Conceptual Model.

The issues raised by submitters can be grouped under four themes:

Compliance

There was both general support for the collection of data and concern around what will be expected of organisations if they do not have the data or do not have it in a form compatible with the draft Conceptual Model. Ensuring the model is sufficiently flexible to avoid creating costs in the future is important.

- Gathering only the data that is necessary is one aspect of this theme. Ensuring that each element included is there for a clear reason is important. In the words of one submitter *"we recommend that MBIE reviews all of the proposed data elements… and include only those that …have a clear purpose, align with the NSSI Vision…, will result in action, and the benefits associated with [analysing] the data outweigh the costs of collecting and managing the data".*

  One submitter observed ".. *that the set of data collected needs to be as complete as possible to be reliable, and particularly is not limited to traditional "academic" measures that are easy to collect, but the collection cannot impose too large a compliance burden on research institutions….. The amount and type of data collected needs to reflect the size and type of the project."*

  Another noted that in designing a system *"…we must strive to minimise complexity and compliance activity/cost in this exercise while recognising the value that is sought by working towards an integrated data system."*

  "…the value proposition/benefit to research providers is not obvious – potentially viewed as another hurdle… on the way to funding…" cautioned one submitter, while another emphasised the need to ensure *"the cost of collection and development of the system does not exceed value of the exercise"* and that *"given the tightness of science institution budgets, there can be a mechanism for funding the system's introduction…".*

- Ensuring that there is future flexibility as opposed to embedding existing institutional arrangements is another aspect of this theme.

  ".. the research, science and innovation system is complex and evolving so we need to be sure that the data collection can encompass outlier cases that don't quite fit the model and also that it can change over time to accommodate new ways of working."

  ".....have you future proofed what info you need to collect if the funding of R&D changes in NZ"?

One example of this is the question of whether there is too much emphasis on *existing* institutional specifics (eg Centres of Research Excellence, National Science Challenges) as opposed to their function of 'strategic alignment'. In one submitter's view, the model appears *"…designed for the present. CoREs, NSCs and concepts of benefitting region, may all date the model. In our view, there will be an ongoing need for a process of review to both deprecate legacy areas and terms, and to allow adoption/incorporation of future programmes."*

Another submitter *"…would recommend abstracting away from the [NSC & CoRE] naming ideas. [These are] ….. open to change. For longevity ...recommend calling the concept something like "Strategic Government Alignment" (or similar)".*

These aspects were reflected in both general comments and in specific comments on data entities and elements.

Comprehensiveness

When asked how well the model represents the RS&I system at a high level, the overall response was that most elements of the system are represented but that further work is needed on both definitions and the relationships between elements.

Submitters recognised the strengths of the initial draft and identified a number of areas where they believed further work could be helpful:

- The potential scope of the NRIS is very broad[5]. Ensuring the model is flexible enough to capture the full extent of the RS&I system is a key aspect of this theme. One submitter observed that *"…The box diagram on page 13 [of the consultation document] is very nice and represents how it works sometimes, but in many, many cases it is not how it works."* While *"the model maps well the flows of activities and outputs that follow research investments or `awards'…", "…not all outputs and collaborations result from*

[5] See page 8 of discussion document (http://www.mbie.govt.nz/info-services/science-innovation/research-and-data/pdf-library/rsi-data-conceptual-model-consultation-draft-march-2017.pdf):
"This model is intended to apply broadly to:
1. All RS&I activities funded in whole or in part by the New Zealand government
2. All RS&I activities performed in New Zealand state sector organisations, such as Crown Research Institutes, universities and Callaghan Innovation.
3. Other organisations may wish to adopt the model and associated definitions and elements on a voluntary basis."

the allocation of specific research investments…" and *"Not all research is funded via an application. Many Outputs are produced through unfunded research."*

The model represents the system *"Fairly well…"* but *"…omits any internally funded research (e.g. university internal research …)…"* and assumes *"…an idealised, linear research/funding pathway [when] research and innovation does not always progress in this way….. It is often complicated to understand the relationships between funding, projects, outputs and outcomes."*

- The very wide range of funding pathways and the varied nature of collaboration, especially within large institutions and between the public and private sectors, was another area that attracted comment from several submitters. One submitter observed that they have a *"many to many relationship in some cases between funding (grants) and projects, i.e. one grant can be linked to more than one project, and a project could have more than one funder. We do not see this relationship as a lead funder and co-funder, rather both as lead funders."*

  Specific examples included questions around the treatment of in-kind support, sub-contractors and international support and partnerships within the model.

  Another submitter observed that the *"proposed RS&I data model appears to stem from an overly simplistic view of the research endeavour. For example, it would ask organisations to link outputs such as spin-off companies, licences, products, inventions, standards or policies, and research techniques to particular grants or awards. In reality the research process is much messier than this – it is seldom this linear, with many contributors from a range of organisations and countries, especially for true breakthroughs or new technologies."*

  A further submitter noted that in their view *"…the model does not capture the engagement between applicant and end users (or perhaps 'next users' would be a better term) before the applicant submits an application. Similarly, the funder would typically engage with end users before setting up a fund. I think this is reasonably important. The so-called triple-helix model of universities emphasises the value of pre-research engagement between researchers, industry and government. It is not helpful*

to conceptualise end users as collaborators with research and receivers of outputs; they have a crucial role in research problem definition."

- Several submitters sought greater clarity around the way commercial research and commercialisation of research was addressed by the model. Several submitters noted that, despite the scope statement *"The concepts in the document are very much oriented toward government science investment rather than taking in the whole innovation system".* Another submitter observed that *"This conceptual model does not fit well for commercially funded R&D scenarios…"*.

- Whether to include proposals whose funding is declined was a topic that attracted some comment. One submitter observed that *"…there needs to be a focus on un-funded proposals also, especially around reasons for not funding. The significant majority of proposal-writing is unsuccessful and represents a major loss to the sector."* while another remarked that *"We only keep information on successful [funding applications]…"*

There are a number of detailed comments on the Funder/Award/Application/Co-funder section and around issues related to Outputs/End-Users and collaboration. Together these raise questions around the model's comprehensiveness. Appendix 3 sets out comments on each of these data entities.

Comparability and Compatibility

Submitters support the *"the principle of using recognised data standards and unique identifiers".* ORCID in particular attracted favourable comment from several submitters .Generally, submitters were keen to ensure consistent definitions and alignment across different agencies and alignment with international definitions.

- Getting definitions clear and making good use of international standards is supported, and the effort that had gone into this is recognised. Some questions however remained. At the conceptual level, one submitter observed that *"…a definition of what is covered by RS&I would be helpful."* while another asked *"What is the difference between the D in R&D and 'non-science innovation'? R&D seems to be a stand-in for just "R" or "science" these days. It's a nomenclature legacy issue, which is increasingly incommensurate with multiple other definitions in current usage (including TEC's understanding of terms)"*.

  Another submitter asked *"How does the Conceptual Model presented compare to international best practice? Does an existing system in another small advanced economy exist? E.g. Singapore, Israel, Finland, Ireland etc? If yes, what has been learnt from this? Were there any synergies and learnings at a systems level; regarding how science information was captured and reported at a government level? If there is no international best practice to learn from then this is an ambitious undertaking by MBIE."*

- Alignment with other projects rather than spelling out another set of data requirements is another important aspect of this theme. There was strong support for the ORCID iD project across several submissions and one was *"…at pains to emphasize their support for the use of ORCID iDs as the researcher identifier of preference in New Zealand."*

- The general issue of data integration through the use of mandatory data elements was also raised by some submitters, with one noting that *"…a stronger emphasis on standardisation (using mandatory core variables) would allow for more effective integration".*

- More generally, one submitter asked that *"Where categories in the conceptual model and code sets are more granular than existing international standards, efforts should be made to ensure they are treated as subcategories within the more widely recognized frameworks to facilitate international comparisons when 'rolled up'."*

- A particular aspect of this is whether reductions in the transaction and compliance costs can be achieved if different parts of government require different ways of coding the same information. Several submitters noted the challenge of having *"…two different Government-mandated publication classifications – the PBRF categories, and the ones in this proposed data model. There should be only one."* Problems arise where NRIS standards are *"… contrary to the standards used in other government systems (e.g., the definition of outputs c.f., the PBRF). The true value of this model will come with the alignment of other processes to the model (or the alignment of this model with other processes), and the integration of standards/identifiers into one consistent framework."*

  The Statistics NZ R&D Survey was also raised by two submitters, one of whom noted that the survey *"… also requires data on: external research income by source, external use of research outputs, staff FTE and headcount involved in research, identification of sectors that our research relates to, and type of research undertaken (basic, targeted, applied, experimental development)."* but the *"…definition of research …[in the survey]… is similar (but not identical) to the definition in the proposed RS&I data model."*

- A related concern at the code set level is whether the coding aims for more precision and detail than can be achieved at present. As one submitter observed: *"The risk in keeping the request at such a disaggregated level will be that the NRIS will lose a significant portion of information (all instances where the information is available at an overarching category level but not at a drilled down level)*".

Confidentiality

Submitters want to know how the NRIS will deal with confidentiality of both commercial and personal information.

- A common area of comment was the apparent need to supply sensitive co-funding or client information especially where the research involves commercial partners.

One submitter was "*very concerned about the proposed plan for the Government to require research organisations to supply data on private funders and private funding, including where the funding has no connection to Government research funding."* Others noted that providing such data may *"…breach confidentiality agreements."*

The sensitivity of commercial information was also mentioned: *"Even the fact that [a] company makes [an] application can be commercially sensitive and allow inferences about their commercial strategy"*. Indeed "*details of the funder and the titles of the grant, project or outputs would be sufficient to reveal information that would be in breach of the funding contract and would jeopardise both current and future funding from the organisation if made public".*

Submitters stressed the need to ensure *"…information on commercially sensitive research is thoroughly protected."* Concern was expressed that *"The model makes no allowance for legally privileged information, such as where commercial sensitivities or confidentiality agreements exist."*

- Concerns were also expressed that seeking confidential *personal* information from private sector partners could result in reduced interest in such partnerships.

  "Capturing sensitive personal information about staff in external businesses (ethnicity, qualifications etc.) would be inappropriate …. and if required would likely deter some businesses from working with us".

- Several submitters emphasised the need to ensure the privacy of individuals is protected, including in the review process.

  "Reviewer information as proposed would likely be a breach of promised confidentiality, and a waiver required of reviewers can be expected to impair the ability to get reviewers willing to offer frank reviews…" The importance of collecting this information (while ensuring it is *"suitably anonymised"*) was also noted: *"Reviewer and Application data must be mandatory. These are processes designed to make decisions around millions of dollars of taxpayer funds and if we are to have*

accountability, these processes must be able to provide their decision-making information".

The value of anonymised demographic information also attracted comment: *"Researcher Demographic data must be mandatory. If we are to understand our research community we need better information now…".*

What parts of the RS&I Data Conceptual Model were the main focus of detailed comments?

Detailed feedback on elements of the model is somewhat concentrated around the funding and funding/project nexus and the collaborator/stakeholder/end-user nexus. Output definitions are also important to submitters, especially those currently providing output information using PBRF codes.

Comments on the Funder/Fund/Applicant/Recipient entities were closely related and often highlighted specific data or classification issues.

The Primary Award entity received comments from many submitters, with related comments on the Co-Funder and Co-Funding entities. Many submitters made similar points.

A large number of very specific comments were received on the Project and Researcher entities, again many were similar and related to data and classification issues.

By contrast, the Infrastructure, Reviewer and Application entities received relatively limited comment at a detail level.

Comments for each entity are included in Appendix 2. These are best read in conjunction with the high level themes set out earlier, as the themes provide context for some specific comments.

Appendix 1: Summary of Individual Submissions by Submitter

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
Auckland University	Auckland University supports NRIS but they have several concerns. First, they believe that providing such detailed information would increase transaction cost and decrease research productivity. They would like to note that many of the proposed data elements could not be provided using the current system. Even though they are supportive of a staged roll-out, they would like to have funds available for developing a new research management system as well as covering ongoing cost for integrating current data into NRIS. In addition, they are concerned about providing	Auckland University believes that the model does not capture the complexity of the research process and especially the challenge of linking research outputs to grants. From their perspective the model is too detailed and loses critical focus on the vision and purpose. They provide a set of criteria for selecting standards and data elements and would like MBIE to take them into account.	Auckland University suggest that the data elements for entities 'Funder', 'Fund', 'Applicant', 'Primary Award', 'Researchers', 'End user collaborators', and 'Outputs' should be assessed carefully to limit the mandatory requirements.	MBIE should clarify whether government will have nationwide code sets for outputs type and organisation type that will be used for PBRF, R&D Survey and NRIS.

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	information on private funders and negative consequences they may face as a result. Furthermore, they suggest collecting data from original sources wherever possible. They believe that the feeds into NRIS should be biannual instead of live or monthly. Finally, they suggest that providing duplicate data to multiple government sources should be avoided.			
AUT	AUT supports the use of unique identifiers for researchers and institutions. AUT is concerned about the administrative burden of data collection and confidentiality of private co-funders. AUT would like MBIE to clarify the drivers and priorities and specify the benefits of this project to	AUT believes that the conceptual model captures the essentials elements of external funding but not in-kind. They believe that there are too many optional categories and they are unclear whether this undermines the model. They believe that the definitions for Primary Award and Collaboration are not particularly explicit	AUT believes that the data entities 'Project', 'Primary Award', 'Project', 'Researchers', 'End User Collaborators' and 'Outputs' need more clarification.	AUT believe that 'Co-funding Type' and 'Personnel Roles' code sets are missing some codes.

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	organisations. They also suggest that MBIE should clearly state who provides data for each key elements. They also would like to know how the 'Conceptual Model' presented compare to international best practices.	or clear.		
Business New Zealand	Business NZ believes it is important that the Government can assess value from CRI's and the tertiary education system. Good information can help clarify where New Zealand gets the best return on its R&D investment.	Business NZ thought the model looked comprehensive. They also asked whether MBIE is confident the model is future-proofed in terms of possible changes to the research and development funding system.		
Callaghan Innovation	Callaghan Innovation supports NRIS. However, they have concerns with the collection and the provision of commercially sensitive company data into (NRIS). In addition, they believe	Callaghan Innovation thinks that the model is a fair representation of the core entities and concepts of the research, science and innovation system. However, they believe that it does not fit well for	Callaghan Innovation would like MBIE to further refine the current definitions of 'Researcher' and 'Output' concepts.	Callaghan Innovation suggests updates of 'Code Set Fund Type' and 'Code Set Project Type' to cover all of their activities by the proposed codes.

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	that compliance burden in terms of system changes and costs will likely outweigh the benefits. They suggest commercially funded R&D should be excluded. Callaghan Innovation believes that requiring NZBN and ORCID from their customers will make it hard for businesses to work with government and deter potential customers from engaging with them.	commercially funded R&D scenarios. In addition, they suggest that 'Outputs' should be linked to 'Project' entity.		
Digital Science	Digital Science is supportive of NRIS and using ORCID across the system.	Digital Science believes that the high-level conceptual model of New Zealand's research and innovation system is very well represented.	While Digital Science believes that the core data elements and entities are very detailed, they believe that outputs can be differentiated to apply different models for e.g. publications, patents, and clinical trials. In addition, they would like MBIE to explain which organisation will be providing the information on each entity to the	Digital Science suggests separation of organisational structure and research themes for the code sets.

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
			system. They also recommend tracking funding by year.	
Health Research Council	The HRC is of the view that quality trumps quantity and therefore that it is more important to have data of a high quality on a smaller number of critical data elements, than data of dubious quality on a larger number of elements, containing not only critical but also 'nice to have' information.		HRC believes that the data entities 'Funder', 'Primary Award', 'Projects', 'Researchers', 'Outputs' and 'End User Collaboration' need more clarification.	HRC emphasises that some of the codes within the 'Personnel Role' element and within the 'Outputs' entities are overlapping. Also, they suggest MBIE consider merging some of the codes to avoid losing information. Finally, they note that the code set for 'Output Type' does not include a code for 'drug development'.
Herve Thevenon	Herve Thevenon believes that the conceptual model needs to be accompanied by information on processes, and is concerned this is absent from the consultation document. He believes the lack of past, present and future processes puts at risk the sustainability of the model	Herve Thevenon would like to see • all terms used in the model clearly defined; and • all definitions presented before diagrams of the model.	Herve Thevenon believes clear definitions for 'National Infrastructure', 'Outputs', 'Research Community' and 'External Partner' are missing from the document. He also believes that without clear agreement on the higher level of the model, one cannot discuss the suitability of data	

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	and reusability of data. He would like the conceptual model amended so it has the ability to incorporate private research, which operates using different processes.	He suggests removing the 'Start" and "End" boxes as they are not entities, and reiterates that it is hard to comment without process information.	elements and entities.	
KiwiNet	KiwiNet is pleased to see the commercialisation is identified to be one of the core concepts but there is a concern that research commercialisation aspect of the model is not well developed. They would like it recognised as a separate downstream activity in the model.	KiwiNet believes that model must be enhanced to capture inputs and outputs from downstream research commercialisation activity, particularly that funded by MBIE through the PreSeed Accelerator Fund. Outcomes such as start-up company formation and licence deals, result from the commercialisation process and should not be defined as research activity outcomes in the model.	'Products' and 'Patents' should be removed as research outcomes, and included as commercialisation outcomes. Initial research 'outcomes' that would form 'inputs' to the commercialisation process could include a 'disclosure' to a university Technology Transfer Office or CRI commercial uni.t	The 'License' and 'Spin Off Company' code set outputs should be captured as 'commercialisation' outputs
Landcare Research	Landcare appreciates the acknowledgement that adaptation and implementation will take	Landcare wants MBIE to clarify the distinction between an 'Award' and a 'Directly Awarded Project'	The data entities Applicant, Primary Award, Project, Recipients, Researchers, End User Collaborators and	Landcare would like MBIE to ensure that terminology for 'Personnel Roles' is consistent with the ones

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	some time. Landcare is unclear how the actual ORCID system fits in and what kind of data can be retrieved from ORCID to populate NRIS. Landcare is concerned about providing commercially sensitive data.	as they assume that reporting only applies to the former. They believe that the definition of Collaboration is not clear and is open to interpretation.	Outputs need more clarification. Landcare needs clarification on the role of 'Recipients' and 'Research Collaborators'.	specified in the fund applications. Also, some of the definitions for 'Personnel Roles' are no longer stated in the Endeavour fund guidelines. Landcare believes that 'Co-funding' type set could be simplied.
Lincoln University	Lincoln University is supportive of system-wide unique identifiers for both researchers and institutions but notes that developing researcher ORCID iDs takes time and that it would be desirable for funders within NZ to all request a single researcher ID type in funding applications.	Lincoln University believes that the model does not represent the RS&I system of New Zealand as from their perspective it doesn't capture the pre-research engagements and the role of end users in the system. In addition, they believe that it misses the feedback between end users' need and funders' intentions.	Overall, they believe that the data elements and entities are comprehensive. Additional fields are proposed for the Project element. In addition, they would like MBIE to provide a guideline on how to evaluate 'researchers' career stage'.	
Marsden Fund Council	The Marsden Fund Council is supportive of the initiative to set common data standards and allow for a system-wide view and assessment of the New			

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	Zealand research sector. The Marsden Fund Council acknowledges — and endorses — the submission by the Royal Society Te Apārangi on the draft Research, Science and Innovation Data Conceptual Model.			
Massey University	Massey University has concerns on providing confidential data since that may jeopardise both current and future funding from organisations if made public.	Massey University believes that the model does not capture many-many relationship between funding and projects.		
Ministry for Primary Industries	MPI is concerned over commercial sensitivity. In addition, MPI is concerned about supplying data about unsuccessful applications, due to commercial sensitivity, and lack of perceived value for that data.	MPI would like MBIE to clarify how the model differentiates between a programme and a project. MPI's fisheries research includes cost recovery funds; MBIE needs to clarify how this will be handled in the model.	The data entities 'Funder', 'Fund', 'Applicant', 'Application', 'Reviewer', 'Primary Award', 'Co-funder', 'Projects', 'Recipients', 'Researchers' and 'Outputs' need more clarification.	

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
Ministry for the Environment	MFE in general supports the objectives and endeavours of this project. MfE suggests that the detailed schematic representation of the model should be edited for clarity and simplicity. MfE asks MBIE to clarify whether all of the required information is 'critical'.		MfE would like MBIE to clarify whether End User Collaborators can be individuals and how Outputs and Projects are linked in the model.	MfE needs clarification on how intangible outputs are dealt with.
NZ Association of Scientists	NZAS supports improved data and metrics surrounding research funding if they are openly available in anonymised form. They suggest that un-funded proposals should also be captured by the model. They also support the use of unique identifiers. They think the schematic of the diagram should be updated in a way that the size of the boxes relates to the importance of the entities.	NZAS believe that the model represents how the research, science and innovation system works sometimes but does not apply to many, many cases. Also, they believe that it is not clear who the stakeholders and end users are. In addition, they believe that the model does not capture stakeholder support.	NZAS believes that too many fields in the model are optional and suggests 'Researcher', 'Reviewer', and 'Application' related data to be mandatory.	NZAS has concerns around having out of date or wrongly balanced code sets.

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
Office of the Prime Minister's Chief Science Advisor	OPMCSA believes a clear value statement making the case for the overall exercise is needed. Categories in the conceptual model and code sets may in some cases be more granular than existing international standards; ensure that when data is rolled up it is comparable to internationally available data. OPMCSA suggests using consistent language (e.g. distinguish the 'D' in R&D from 'non-science-based innovation', if it is distinct).	OPMCSA believes the model may not fully capture information on: • private or international funders/funding • unsuccessful applications • the share of a project's work attributable to each researcher or co-author.	OPMCSA sees a need to clarify definitions of 'Applicant', 'End user', 'Output' and 'Commercialisation'. OPMCSA would like MBIE to explain how 'Applicant' differs from/relates to a 'Researcher' and a 'Funder' (e.g. how to distinguish in the model when the applicant is a platform group or institution).	OPMCSA believes some code sets for service outputs are not clear. There are some output types that are not captured in the code sets (e.g. new classification for astronomical or biological entity, mathematical proof, segment of code, practice guidelines, and new methods and processes placed in public domain and not patented).
Plant and Food Research	Plant & Food supports the intent of the NRIS initiative. They see the use of ORCID and NZBN as positive but are unclear what might be used for international organisations. They would like to point out that setting identifiers by the lead organisation	Overall Plant & Food believes that the model maps well the flows of activities and outputs that follow research investments/awards. However, they think the model doesn't capture the cases where the outputs and collaborations are the	Plant & Food believes that some of the data elements that are regularly asked by the funders are not captured in the model. In addition, they query the need to align outputs with ANZSRC classifications since the award that led to the outputs is already	They suggest MBIE include 'All NZ' and 'International' in the code set for Benefiting region.

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	may be incompatible with the systems used by the subcontracted organisations or the trail of association can be lost. In addition, they believe it should be clarified how international funders or subcontractors are represented in the model. They also suggest that ORCID iD should be used across all government funders and be a mandatory field for researchers.	outcome of multiple funding sources in combination. In addition, they believe that MBIE should leave out some of the more complex, costly or time-consuming elements from the model.	associated with an ANZSRC code.	
Royal Society Te Apārangi	Royal Society Te Apārangi supports NRIS. They suggest that the schematic representation of the model should be edited for clarity and simplicity. The Society emphasises their support for the use of ORCID iDs as the researcher identifier of preference in New Zealand.	Royal Society Te Apārangi believes that the model overall does a good job of representing the system but will struggle with national collaborations – NSCs and CoREs. They suggest that MBIE should add 'Secondary Fund' as an entity. In addition, they believe	Royal Society Te Apārangi believes that the definitions for the data entities 'Applicant', 'Primary Award', 'Project', 'Researchers', 'End User Collaborators' and 'Outputs' need more clarification. They suggest additional data elements for 'Output'	Royal Society Te Apārangi points out that the Society is missing from the organisation type.

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	They suggest MBIE refer to the Domain Plan to clearly explain why certain data entities and elements are required. They suggest areas where the Conceptual Model would benefit from close engagement with the research sector.	that the model avoids funding attribution. There is also a concern that the model will not be flexible enough to reflect the changes happen over time. While they are supportive of using NZBN to be the unique identifier for applicants, they would like to know how the model accounts for the case where applicants are international organisations or individuals who do not have NZBN.	and 'Funding', and 'Reviewer'. They believe that Reviewer information as proposed would likely to be a breach of promised confidentiality.	
Science NZ	They welcome the implementation of NRIS. There is a concern around the data collection regarding automation, commercially sensitive research, and minimum requirements. They emphasize that the cost of data collection and the development of the system ought not to exceed value	Even though they believe that the model overall represents the RS&I system of New Zealand, they emphasize the need to account for outlier cases and making the system flexible to change over time.		

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	of data captures. They would like to know whether there will be a funding mechanism for the system's introduction.			
Statistics NZ	Statistics NZ is supportive of NRIS. There is a concern that it is limited in terms of providing comprehensive data for macroeconomic statistics since it only captures government funded research activities. Statistics NZ is supportive of unique identifiers for researchers and organisations and the use of ANZSRC. However, they believe that careful considerations should be given to where they are applicable and to the impacts of revisions of these classifications to the system.	Statistics NZ notes that the model identifies both economic and social outcomes that could result from research, science and innovation. They would like to know how the model will be evaluated to ensure that it is working well. They would like MBIE to provide clear definitions of 'End User', 'End User Collaborator' and 'Stakeholder'. In addition, they suggest MBIE list the infrastructures covered in to improve the 'The National Infrastructure' definition. Statistics NZ suggests including 'Communication' as a key concept and	Statistics NZ believes that too many data elements are optional. They ask MBIE to make personal data elements for Researchers mandatory to align with work being undertaken to standardise personal data collection across government. They note that even though some of this information may already be collected in ORCID iD, it is optional rather than mandatory. They emphasize the importance of using the right classification for the right purpose.	Statistics NZ would like MBIE to rework/redevelop 'Organisation Type' and 'National Science Challenge Theme' code sets to ensure that they are reflecting best practice principles.

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
		provides an updated schematic for the model.		
University of Canterbury	University of Canterbury likes the idea of NRIS and can see the potential of a more rigorous evidence base. They are concerned about the burden of implementation, and keen to discuss further the how and when of implementation. They support the use of data standards and unique identifiers and would like to note that for consistency these should be integrated into one consistent framework that all government agencies can refer to. Furthermore, they believe that parts of reporting should be required from 'research funders' rather than 'research providers'. Supplying information on private funding, and	University of Canterbury believes that the model does well under the assumption of an idealised, linear research/funding pathway but would like to point out that the research and innovation does not always progress in this way. Also, they believe that the model doesn't cover in-kind funding, and that reporting of research programmes which don't carry full overheads (including Government department research funding outside of MBIE and Royal Society) would be difficult.		

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	additional data on government funded projects/researchers would increase the transaction costs. Also, with their current information systems they cannot provide all the data, as currently proposed, and believe that appropriate investigations should be undertaken to determine the ease, or not, of research providers in submitting into a new research reporting system. They would like MBIE to explain how the proposed reporting system would deal with privately funded research under NZ privacy law.			
University of Canterbury Library	University of Canterbury Library supports the implementation of NRIS and encourages creating a common language across the sector. They understand that initially	They believe that the document acknowledges unfunded research in its definitions of concepts but the conceptual model doesn't capture the un-funded research. They		

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	required amount of data could place a burden on institutions, funders and researchers. Also, they are aware of the risk of duplication and over complication across the sector and they believe that in time these issues would be resolved.	point out that not all research is funded via an application and many outputs are produced through unfunded research.		
University of Otago	Otago University believes that a nationwide data repository would be useful and MBIE should examine leading/coordinating an approach for selecting a nationwide RIMS. They have a concern about the potential for excessive administrative resource being required for information collation at the expense of resource for research. They suggest that there should be extensive 'beta'	Otago University believes that MBIE should clarify meaning of 'core concept' and 'data element' and ensure terms and labels are used consistently.	Otago University suggests MBIE consider non-traditional outputs, including those from the fine arts (music, drama, theatre studies etc) and items such as datasets when developing list of outputs	Otago University suggests using OCLC document 'Addressing the Challenges with Organisational Identifiers and ISNI (2016)' for identifying code sets.

Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	testing and that MBIE incorporate pilot data collection exercises across different research providers before the full implementation. They suggest some changes to the schematics of conceptual model. They would like MBIE to consider the cases where a person or an institution does not have the required unique identifier.			
University of Waikato	They are supportive of NRIS and acknowledge that only government funded research, science and innovation activities will be captured by the system. They point out that while data volume may appear large, collecting it is valuable for internal			

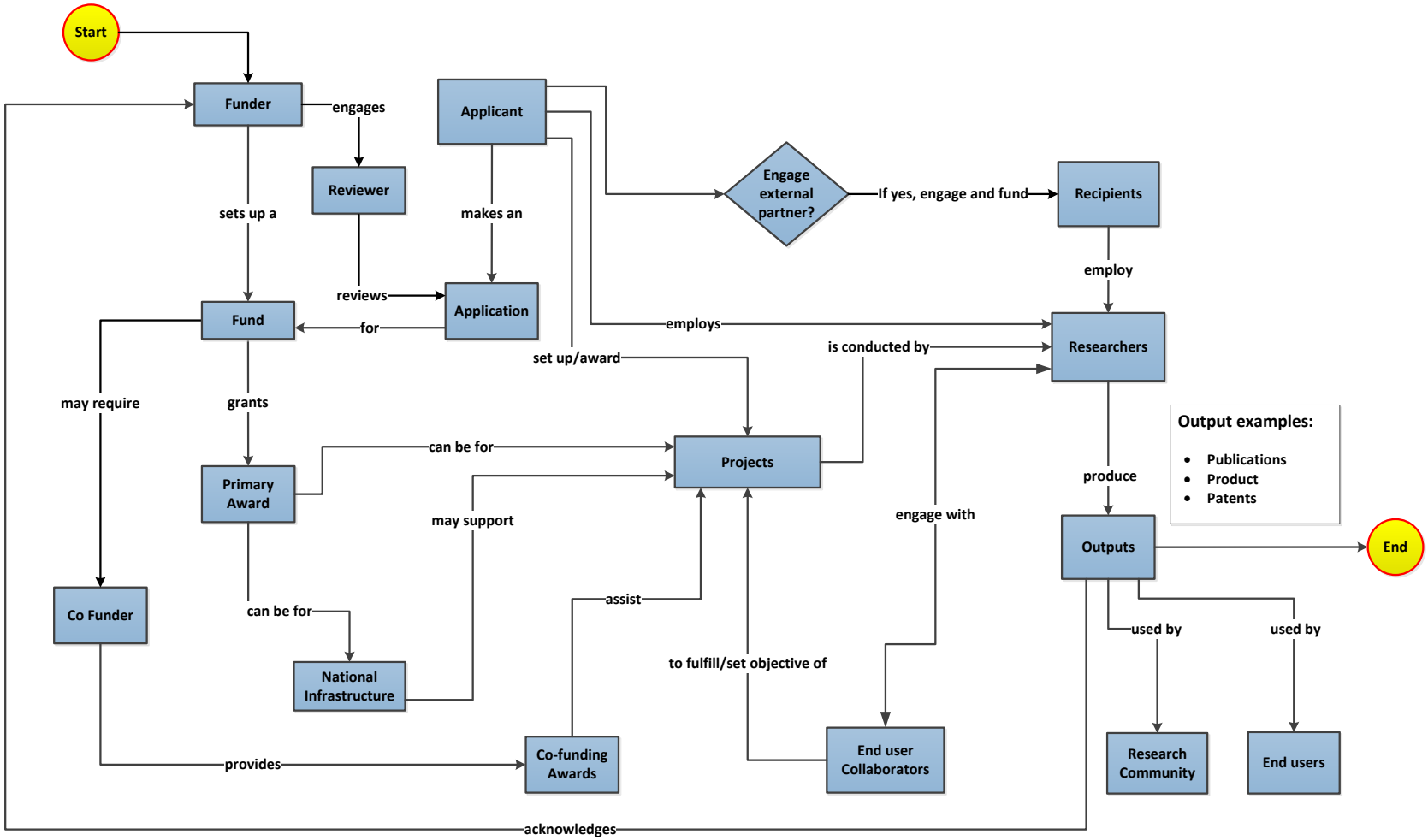
Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
	purposes. They have some concerns that providing information on private companies could affect the willingness of those companies to provide funding to research organisations.			
Victoria University of Wellington	Victoria University of Wellington likes the idea of NRIS. However, they have some concerns over the burden of collecting data regarding automation and minimal requirement and the frequency of feeding data into NRIS.	VUW believes that the model overall captures the RS&I system of New Zealand and includes all critical elements. VUW notes there are commercial sensitivity issues with some forms of co-funding, and would also like to see the term 'co-funding' clearly defined as third party funding (rather than funding from the research provider itself) in the conceptual model. VUW states that the links between research outputs	VUW states that it is difficult to distinguish mandatory or conditional mandatory data elements. They would like MBIE to clarify under which conditions report of conditionally mandatory data elements would be required. They note that there are some additional concepts (e.g. career stage) that still need to be defined.	VUW points out that there should only be one nationwide publication classification. Also, they do not view some of the categories in the data model as research outputs, but as other kinds of research activities.

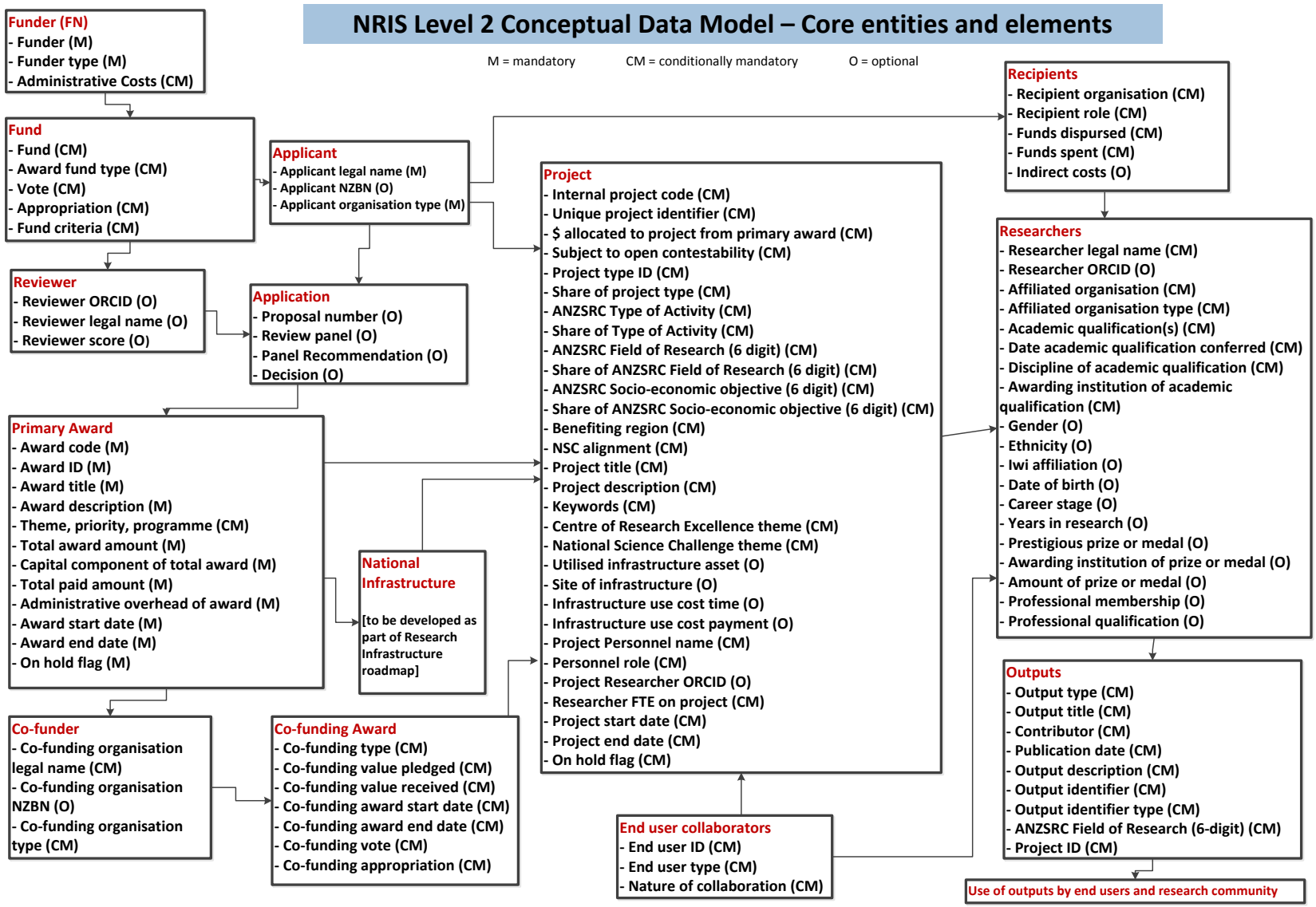
Submitter	General Comments	Core Model	Data Entities and Elements	Code Sets
		and projects are often not sufficiently linear to be able to easily or accurately be apportioned as a percentage contribution of a particular fund/project. Requiring this information could add significant workload to research reporting since only researchers can provide accurate information.		

Appendix 2: Diagrams from Draft Conceptual Model Consultation Document

The following two pages contain the Level 1 and Level 2 Draft Conceptual Model diagrams that were contained in the consultation document released in March 2016. These diagrams are included for ease of reference in relation to comments in Appendices 1 and 3.

NRIS Level 1 Conceptual Data Model – Core entities and concepts




NRIS Level 2 Conceptual Data Model – Core entities and elements
M = mandatory
CM = conditionally mandatory
O = optional
Funder (FN)
- Funder (M)
- Funder type (M)
- Administrative Costs (CM)
Fund
- Fund (CM)
- Award fund type (CM)
- Vote (CM)
- Appropriation (CM)
- Fund criteria (CM)
Applicant
- Applicant legal name (M)
- Applicant NZBN (O)
- Applicant organisation type (M)
Reviewer
- Reviewer ORCID (O)
- Reviewer legal name (O)
- Reviewer score (O)
Application
- Proposal number (O)
- Review panel (O)
- Panel Recommendation (O)
- Decision (O)
Primary Award
- Award code (M)
- Award ID (M)
- Award title (M)
- Award description (M)
- Theme, priority, programme (CM)
- Total award amount (M)
- Capital component of total award (M)
- Total paid amount (M)
- Administrative overhead of award (M)
- Award start date (M)
- Award end date (M)
- On hold flag (M)
National Infrastructure
[to be developed as part of Research Infrastructure roadmap]
Co-funder
- Co-funding organisation legal name (CM)
- Co-funding organisation NZBN (O)
- Co-funding organisation type (CM)
Co-funding Award
- Co-funding type (CM)
- Co-funding value pledged (CM)
- Co-funding value received (CM)
- Co-funding award start date (CM)
- Co-funding award end date (CM)
- Co-funding vote (CM)
- Co-funding appropriation (CM)
Project
- Internal project code (CM)
- Unique project identifier (CM)
- $ allocated to project from primary award (CM)
- Subject to open contestability (CM)
- Project type ID (CM)
- Share of project type (CM)
- ANZSRC Type of Activity (CM)
- Share of Type of Activity (CM)
- ANZSRC Field of Research (6 digit) (CM)
- Share of ANZSRC Field of Research (6 digit) (CM)
- ANZSRC Socio-economic objective (6 digit) (CM)
- Share of ANZSRC Socio-economic objective (6 digit) (CM)
- Benefiting region (CM)
- NSC alignment (CM)
- Project title (CM)
- Project description (CM)
- Keywords (CM)
- Centre of Research Excellence theme (CM)
- National Science Challenge theme (CM)
- Utilised infrastructure asset (O)
- Site of infrastructure (O)
- Infrastructure use cost time (O)
- Infrastructure use cost payment (O)
- Project Personnel name (CM)
- Personnel role (CM)
- Project Researcher ORCID (O)
- Researcher FTE on project (CM)
- Project start date (CM)
- Project end date (CM)
- On hold flag (CM)
End user collaborators
- End user ID (CM)
- End user type (CM)
- Nature of collaboration (CM)
Recipients
- Recipient organisation (CM)
- Recipient role (CM)
- Funds dispursed (CM)
- Funds spent (CM)
- Indirect costs (O)
Researchers
- Researcher legal name (CM)
- Researcher ORCID (O)
- Affiliated organisation (CM)
- Affiliated organisation type (CM)
- Academic qualification(s) (CM)
- Date academic qualification conferred (CM)
- Discipline of academic qualification (CM)
- Awarding institution of academic qualification (CM)
- Gender (O)
- Ethnicity (O)
- Iwi affiliation (O)
- Date of birth (O)
- Career stage (O)
- Years in research (O)
- Prestigious prize or medal (O)
- Awarding institution of prize or medal (O)
- Amount of prize or medal (O)
- Professional membership (O)
- Professional qualification (O)
Outputs
- Output type (CM)
- Output title (CM)
- Contributor (CM)
- Publication date (CM)
- Output description (CM)
- Output identifier (CM)
- Output identifier type (CM)
- ANZSRC Field of Research (6-digit) (CM)
- Project ID (CM)
Use of outputs by end users and research community

Appendix 3: Comments on Data Entities and Elements by Entity

This Appendix contains comments from submissions on data entities and elements in the draft conceptual model. Comments are verbatim except as needed to ensure anonymity and clarity.

Funder

General Comments:

- Where do private and international funders/funding get caught in this model? It doesn't look although there is a requirement for projects themselves to report on their funding sources so would this fall through the cracks?
- Feedback loop between end-user's needs and funder's intensions is missing in the model.

Data Element	Comment
FN1-Funder	• It should include NZBN • It should be optional to provide funder name for funders with no link to Government funding.
FN2-Funder Type	• It should be optional to provide funder type for funders with no link to Government funding.
FN3-Administrative Costs	• It is time dependent. So to track efficiencies you need to understand how this works across different time periods. • What is FN3? Collecting the funder's administrative costs has no value. Total administrative cost for all parties involved is more useful • Is this the cost to the organisation for administering the specific fund?

Fund

General Comments:
We support these elements being mandatory only for Government funds.

Data Element	Comment
F4–Appropriation	• Appropriations come and go depending on budget, need to make this future proof

Applicant

General Comments:

- As with Fund (2.2, only mandatory for government funds), should 2.3 only be mandatory for Government Fund Applications and thus provided by funding agencies – they report on who they have received applications from, and which ones were successful (or not).
- Where a funder is not a government agency you anticipate information coming from research providers i.e. we would report in relation to applications we submit to private funders, meaning that we would simply report our own legal name, NZBN, and type of organisation. This would seem to be redundant?
- Need more clarification of definition of applicant
- Concern over commercial sensitivity
- We are unlikely to provide data about non-successful applications for at least some of our funds, due to commercial sensitivity, resourcing and lack of value for that data.
- Sometimes we don't have an applicant; we go directly to a supplier.
- There are companies created specifically for a project (PGP) and disbanded later.
- Definition of 'applicant' is not clear when they are distinct from 'researcher' but also seem to be a funder. Thus, is the applicant assumed to be a platform such as NSC and CoRE or can it be an individual (i.e. how are personnel or training awards treated in the model)?
- We question whether applicant details are mandatory in addition to recipient information. Is MBIE proposing to record the applicant details for every funding application made by a NZ research organisation (including applications for international or charitable funding)? We suggest that the benefits of collecting and storing this information is outweighed by the compliance cost for organisations
- Is there any relationship between Applicant and Outputs entities? No connection to output is specified.
- The engagement between applicant and end users (or perhaps 'next users' would be a better term) before the applicant submits an application is not captured.

Data Element	Comment
AO1-Applicant Legal Name	• Need more clarification on applicant legal name –applicants for funds can be individuals, not legal entities (they form legal entities later usually). • It is unclear whether applicants are actually expected to be anything other than organisations; how does a self-employed researcher fit into this.
AO2-Applicant NZBN	• Needs to be conditionally mandatory. Only potential issue is with international companies. • Applicant's ORCID should be conditionally mandatory, i.e. included when available.

Application

General Comments:

- Funding agencies should provide the data as we are not privy to review panel information.
- We only keep information on successful projects, i.e., this is the same as project data
- Our procurement team, that deal with contracted research, etc., consider the proposal to be confidential
- Application data must be mandatory

Data Element	Comment
N2-Review Panel	• Is review panel the name of the separate reviewers?

Reviewer

General Comments:

- Reviewer information as proposed would likely be a breach of promised confidentiality, and a waiver required of reviewers can be expected to impair the ability to get reviewers willing to offer frank reviews.
- We can't currently supply this information as our processes make it confidential and our systems and processed don't support the collection of this data. Clarity is needed around the 'optional' status of this.
- Reviewer data must be mandatory

Data Element	Comment
W1- Reviewer ORCiD	• Reviewer's ORCID ID should be conditionally mandatory, i.e. included when available.

Primary Award[6]

General Comments:

- Some of the information may change over time. Start date, end date, amount awarded, infrastructure use cost. Even the overhead, salary and capital components will shift when the budget forecast is translated into reality.
- Definition is not clear. We assume that this only applies to awards and not to situations where we are directly awarded a project? Some things which are appropriate for (true) awards, are not appropriate at Project level.
- It suggests there are secondary awards.

[6] A number of submissions commented on the complexity of funding systems and whether the notion of a primary award and co-funding reflected the way the system operates.

- The term ‘award’ can be confused by non-experts with other awards. The term ‘grant’ is more common and could be less confusing as could ‘funding agreement’ indicating the contractual arrangement. The terms pre-award, post-award and ‘award’ all seem quite unique to the research intensive community
- How are personnel or training awards treated in the model?
- Recommend tracking funding by year.
- We suggest that this is mandatory only for Government-sponsored research.
- Funds spent element is missing.
- Since PA2 is system wide, suggest swapping definitions for PA1 and PA2
- Dollars and FTEs proposed and actuals – how are these recorded? (NB: for finances, dispursed is used and instead should be disbursed)

Data Element	**Comment**
PA5-Theme, Priority or Programme	• How is this valuable when every funder has different categories? What if we have theme, priority and programme to list? Will we be able to include many entries here? • We have assumed that HRC Research Investment Stream is sufficient for theme. Is this acceptable?
PA6-Total Award Amount	• In some cases our funds are designed to be flexible with the primary award amount being expected to change – can this flexibility be managed? Not the current approach for our procurement – we advertise the award amount as a range and then announce it as a range, which aligns with government guidelines. So we don’t publicly release this number, as then the competing suppliers would see what their competition has been awarded. How will MBIE manage this confidentiality issue? CRIs bid for the same project and we don’t tell them how much exactly we award, just a range. • For PGP, if we try to extract just the R&D work we will be estimating all of these in an arbitrary manner. • How will primary award amount be related to financial year? And how will financial year be defined – business or government financial year? • Will the NRIS collect data about forecast award amounts – as in, the budgeted amount that will be spent in each financial year of the project?
PA7-Capital Component of Total Award	• We note that Society-administered funds cannot typically be used for capital expenditure, but recognise the utility of such an element.
PA8-Total Paid Amount	• Will Total Paid Amount include time allocations so it can be viewed by financial year, or will it always be total paid up to this point?
PA9-Administrative	• This is arbitrary at programme level, and you will be making an assumption that the applicant budget is broken up in this way.

Overhead of Award	○ Is this valuable? ○ This should be optional. ○ What is administration defined as? What does it include? Does it include project management? Who does this cost relate to? • We have assumed that this is the budgeted cost of the Host administering the award. Is this correct? • A workable cross-sector definition of overhead is absolutely required.
PA10-Award Start Date	• For PGP, if we try to extract just the R&D work we will be estimating all of these in an arbitrary manner.
PA11-Award End Date	• For PGP, if we try to extract just the R&D work we will be estimating all of these in an arbitrary manner.
PA12-On Hold Flag	• Instead of "On hold" data element, suggest it would be better to have a status attribute which could include a type of "on hold".

Co-funder

General Comments:

- Time Stamping is missing. Co-funders change and amounts change.
- There is a need for country of origin, and currency, to account for non-NZ co-funding.
- When identifying co-funder organisation types, it is not clear what level research organisation would be required to report at.
- Is it a fund that needs a co-funder or a project that requires a co-funder?
- A project could have more than 1 funder. We do not see this relationship as a lead funder and co-funder.
- Where do private and international funders/funding get caught in this model?

Data Elements	**Comment**
C1-Co-funding Organisation Legal Name	• We have groups that are not legal entities, and it needs to be able to cope with multiples
C2-Co-funding organisation NZBN	• NZBN should be conditionally mandatory

Co-funding Award[7]

General Comments:

- Definition is not clear. Not clear if cofounding of Projects (not awards) is expected to be reported upon. Co-funding is not straightforward for in-kind funding.
- For PGP, the co-funding is dealt with by the company running the PGP. The details of who the co-funders are, what the amount is, what is actually spent and how it changes over time are not currently provided on a regular basis.
- It is not clear from the document which organisation would be expected to provide this data. This is especially the case of in-kind co-funding awards – at present we do not record this information in systems.
- Co-funders may explicitly prohibit sharing of this information.
- We are not sure if separating the co-funding award money from the primary award money is the best way of presenting the system, as both of them can be used for national infrastructure or supporting a project
- "Co-funding Award" should be defined better.
- Co-funding entity should be excluded from the model as it would raise more issues than could be resolved.
- See also comments under Primary Award about changes over time
- See also comments under Primary Award about Dollars and FTEs

National Infrastructure

General Comments:

- Definition is missing.
- Is it necessary to have data related to infrastructure both in the 'Project' table and the 'National Infrastructure' table? (Do they address different types of structures?)

Projects

General Comments:

- Since P2 is system wide, suggest swapping definitions for P1 and P2
- For P6.1, P7.1, and P8.1 We wonder about the validity/usefulness of assigning the % ANZSRC code splits at the 6 digit level, as in reality the 6 digit code areas are not always mutually exclusive aspects of research. We suggest considering applying the % at only the

[7] A number of submissions commented on the complexity of funding systems and whether the notion of a primary award and co-funding reflected the way the system operates.

4 digit code level, and within that simply specifying which of the 6 digit specific elements applied (yes or no, at that level).

- The order here could be more logical, with, for example, the project title being first, followed by the description, start/end date etc. The rationale behind this is that the information that is at the top will be that which people first seek.
- It may be worthwhile concatenating the NSC and CORE theme work into a single entity which asks "what strategic documents is this work aligned to?". This would enable a researcher to log multiple connections to documents like the NSSI, the Conservation and Environment Science Roadmap, the Primary Sector Science Roadmap, etc.
- Clarity needed around programme vs project
- It doesn't appear that any information is currently sought on the share of a project's work attributable to each researcher/author
- CoREs, NSCs and concepts of benefitting region, may all date the model. In our view, there will be an ongoing need for a process of review to both deprecate legacy areas and terms, and to allow adoption/incorporation of future programmes.
- Add 'project status' field and an optional 'extension date' field
- Link between Project and Output is missing.
- Is it necessary to have data related to infrastructure both in the 'Project' table and the 'National Infrastructure' table? (Do they address different types of structures?)
- See also comments under Primary Award about changes over time
- See also comments under Primary Award about Dollars and FTEs

Data Element	**Comment**
P1-Internal Project Code	• Not clear the benefit of a separate CM identity field. P1 should be optional as P2 is a system wide unique identifier that links various data sources. • Where there is no unique funder generated ID to use as P2, a lead organisation can generate a unique P2 ID to be used across collaborating organisations without necessarily creating a further P1 ID?
P2-Unique Project Identifier	• We suggest for the Project ID (P2) may be created by the lead organisation, OR if appropriate a unique ID created by the funding agency may be used. • Since our contracts are not separated into Award + Projects structure we will need some guidance on how we should generate this.
P4-Subject to Open Contestability	• It should be optional
P6-ANZSRC Type of Activity	• It should be optional
P7-ANZSRC Field of Research	• It should be optional

P8-ANZSRC Socio-Economic Objective	• It should be optional
P9-Benefiting Region	• It should be optional • Please clarify/confirm our interpretation, that if work applies to the whole of NZ we do not indicate any specific regions, vs if the work only applies to subset of regions we specify which ones from the list. • Benefiting region code set should include All of NZ as well as International
P10-NSC Alignment	• It should be optional • Does this mean that NSCs will stop reporting on aligned research? If a project can be aligned to >1 NSC, how will you ensure calculation total spend of research of relevance to NSC objectives without double counting? • We would recommend abstracting away from the NSC naming ideas. NSC is a politically derived funding structure that is open to change. For longevity of the feed, I would recommend calling the concept something like "Strategic Government Alignment" (or similar). • Capturing this will involve lots of researcher and challenge administrator time. Is this justified?
P13-Keywords	• It should be optional
P14-Centre of Research Excellence Theme	• This type of name tends to change over time. I would recommend renaming to focus on the purpose rather than the current name
P16-Utilised Infrastructure Asset	• Usage of infrastructure may be difficult to quantify in many cases? • Utilised infrastructure asset and its children elements, there is a need to define national infrastructure to help determine the level of reporting that may be needed to be recorded (e.g., is this at the level of departmental microscopes?) • Is it necessary to have data related to infrastructure both in the 'Project' table and the 'National Infrastructure' table? (Do they address different types of structures?)
P16.3-Infrastructure use cost payment	• This may change over times • It may be difficult to quantify in many cases.
P17-Project Personnel Name	• It should be optional • Seems to be duplication P17/R1 Is this is to link projects and researchers • What is P17 – can the sub categories here have multiple entries? Why list researcher here when it is under researcher?
P17.1-Personnel Role	• Need confirmation of these categories as there are overlaps • It should be optional

P17.2-Project Researcher	• Seems duplication of R2. Is this to link projects to researchers?
P17.3-Researcher FTE on Project	• It should be optional
P18-Project Start Date	• It may change over time
P19-Project End Date	• It may change over time
P20-On Hold Flag	• It should be optional

Recipients

General Comments:
It should be captured from the Recipient organisation not the Applicant organisation.

Data Element	**Comment**
S1-Recipient Organisation	• What is the source of this identifier? It is unclear whether recipients are actually expected to be anything other than organisations; how does a self-employed researcher fit into this system?
S2-Recipient Role	• Do you actually need to know who (for example) we paid to run a few routine lab samples, or who we paid to service some possum traps? If there is real value in knowing this detail then please elucidate, otherwise consider what value is to be gained vs additional administrative burden collating/providing this level of detail.
S3-Funds Disbursed	• S3 defines the Funds Disbursed (which implies paid to date), but then reason refers to funds allocated. We allocate funds through a contract across the full duration of a project subcontract. We periodically (can be several times per year) disburse funds on receipt of invoice and against progress on contracted work. Need clarification • How does it differ from S4?
S4-Funds Spent	• S4 Funds Spent - Expectations around how much financial information a lead contractor can reasonably extract from a subcontractor, and without incurring additional transaction costs, need to be considered alongside issues of commercial sensitivity. • How does it differ from S3?
S5-Indirect Costs	• How much information a lead organisation can reasonably be expected to have on a subcontractor's financial information. • What does this mean? This appears to be a confusing concept – is there a very clear definition of Indirect?

Researchers

General Comment:

- We need clarity around what will be extracted/linked using ORCID, and assurance that we will not be expected to provide information able to be extracted from a researchers ORCID record (e.g. R3-R4.3, and perhaps in future R5-R13 (R5-R8 data does not change – gather once))
- Need more clarification – what is the definition of a researcher? Sometimes, civilians are used to collect data (e.g. Sustainable Farming Fund), are they a researcher?
- Is the ORCID number going to be used in such a way that we won't need to collect researcher identifying details separately?
- Some of the data elements required under Researchers and Output fields point to activity not necessarily associated with single awards but rather is the result of historic career-long activity. This data should be sourced from the ORCID but these data elements are not mandatory in ORCID.
- Will research organisations be feeding data into the NRIS separately?
- What about publication record – is this recorded through ORCID?
- What about subject matter expertise outside of academic qualification?
- Add Maori Descent and Address fields.
- The use of timestamping is suggested for Name, and Address data elements.
- Funders have asked for different researcher IDs from Scopus to ORCID in their funding applications and a decision at a national level would be welcome.
- How will the expertise of NZ researchers be determined?
- Projects and outputs have ANZSRC codes, but not people. How are output ANZSRCs assigned/collected?
- See also comments under Primary Award about Dollars and FTEs

Data Element	Comment
R1-Researcher Legal Name	• Researchers legal name should be specified with three variables: Official Name, Preferred Name and Former Name. Also, first, middle and last name. • Researcher Legal Name, will need to be aware of multi-name individuals, and changes of legal name.
R3.1-Affiliated Organisation Type	• Assume type? Should it be 'The type of organisation(s) with whom a researcher is affiliated'
R4-Academic Qualifications	• It should be recorded using the NZQF • It should be optional.
R4.1-Date Academic Qualification Conferred	• It should be optional.

R4.2-Discipline of Academic Qualification	• Discipline of academic qualification should use NZSCED or ISCED-F • It should be optional.
R4.3-Awarding Institution of Academic Qualification	• It should be optional.
R5-Gender	• It should be mandatory • Instead of Gender, Gender Identity or Sex should be used. • It should be based on Stats NZ Gender Identity Classification
R6-Ethnicity	• It should be mandatory • It should be collected at level 4 of the classification
R7-Iwi Affiliation	• It should be mandatory • Stats NZ standard should be mandated.
R8-Date of Birth	• It should be mandatory • It should comply with the requirements of Stats NZ Statistical Standard for Age.
R9-Career Stage	• Career Stage should be defined clearly. • A guideline on how to calculate the Researcher's individual career stage should be provided. At what point in their career would a researcher move between 'early and middle' career levels? On employment years from the first lecturer/researcher appointment or level of publications/research activity?
R10-Years in Research	• Definition of active in research is not clear
R11-Prestigious prize or medal	• What are the inclusions in conceptual model for prize or medal? • Would you expect report this (prize/medal/memberships/qualifications) only in the year of being awarded, or every time you are on a project x every year? Simpler if this information is gathered once (not for every project).
R11.1 Awarding Institution of Prize or Medal	• Would you expect report this (prize/medal/memberships/qualifications) only in the year of being awarded, or every time you are on a project x every year? Simpler if this information is gathered once (not for every project).
R11.2 Amount of Prize/Medal	• Would you expect report this (prize/medal/memberships/qualifications) only in the year of being awarded, or every time you are on a project x every year? Simpler if this information is gathered once (not for every project).

R12-Professional Membership	• Would you expect report this prize/medal/memberships/qualifications) only in the year of being awarded, or every time you are on a project x every year? Simpler if this information is gathered once (not for every project).
R13-Professional Qualification	• Would you expect report this prize/medal/memberships/qualifications) only in the year of being awarded, or every time you are on a project x every year? Simpler if this information is gathered once (not for every project).

End User Collaboration

General Comments:

- Will this log both the individuals involved as well as where it is an organisation that is collaborating?
- End User Collaborators will require testing with the sector, together with its proposed Collaborator Type code-set.
- The proposal to collect information about end user collaborators seems misguided.
- In many cases there will be confidentiality arrangements in place in relation to many end user collaborators which would prevent this information from being shared.
- Instead of using 'End user Collaborators' as a separate box, we suggest creating a relationship between 'Researchers' and 'End Users' that shows the collaboration relationship.

Data Element	**Comment**
E1-End User ID	• Would not recommend storing different types of information within the same area, split it up to give confidence in the data quality. • Use of the NZBN should take precedence noting that the ANZSIC classification is dated and isn't particularly well suited for identifying and differentiating projects or institutions undertaking research and development i.e. there is only one class for scientific research services in ANZSIC. • This is currently quite vague and broad regarding the unique identifier to be used, it could come from a number of different classifications, (NZBN, Stats NZ Iwi classification, ANZSIC, free text and we aren't clear on how that would work in practice.
E2-End User Type	• We can provide a partial dataset for this field. However, is this data of use without the associated End User ID which we do not have?
E3-Nature of Collaboration	• Can the nature of collaboration (E3) be >1 type?

Outputs

General Comments:

- What is meaningful? Is outcome better than output?
- ISBN and URL would be useful fields in the outputs section
- Where are the social outputs, like community support, decision making tools, etc. – how will outputs deal with these more intangible outputs?[8]
- Recommend differentiating the "outputs" to apply different models for e.g. publications, patents, and clinical trials.
- No quality indicator is specified for outputs.
- We note that goods, services, products and publications are all recognised as outputs; however, the elements are all very publication-centric.
- No elements for output costs is specified
- We are not able to link research outputs to funded projects, and therefore suggest that the project ID be optional.
- Collecting information on outputs- This information would link back to a Project from an outputs repository like Research Elements.
- Some of the data elements required under Researchers and Output fields point to activity not necessarily associated with single awards but rather is the result of historic career-long activity
- There seems to have been over-reliance on traditional outputs such as publications. Consideration needs to be given to non-traditional outputs, including those from the fine arts (music, drama, theatre studies etc.) and items such as datasets.
- Link between Project and Output is missing.

Data Element	Comment
O4-Publication Date	• Traditional research outputs don't encompass everything, for example commercial outputs often are not published in any way, not even through patents. • Make clear that "Publication date" is the creation/opening/etc date element for non-publication outputs.
O6-Output Identifier	• Need clarification – not clear what is required/how the output gets an identifier e.g. is this a DOI or ARK, is it mandatory to give all outputs an identifier such as these, or are more informal identifiers OK (e.g. internal contract report number)?
O7-Output Identifier Type	• Need clarification – not clear what is required/how the output gets an identifier e.g. is this a DOI or ARK, is it mandatory to give all outputs an identifier such as these, or are more informal identifiers OK (e.g. internal contract report number)?

[8] Several submitters queried whether the model adequately captured Humanities and Social Science research in general

O8-ANZSRC Field of Research at the 6-digit level	• It is onerous and unrealistic to expect each and every output to have an ANZSRC code assigned. Outputs should inherit ANZSRC codes from the project (P6-P8.1) from which they are produced - if this is the intention please make that clear. • How will these codes be collected? • *We query the need for ANZSRC codes for outputs directly associated with awards as well as at the Project level; what is this designed to check? Are you concerned that outputs might not align with the intended FOR for the award? This seems onerous and should be optional. • We also do not identify ANZSRC field of research codes for our research outputs and suggest that this be optional. Much of our own data on research outputs is fed from international databases which do not use ANZRC codes and *therefore it would be difficult to add and verify this information.*
O9-Project ID	• Presumably this is the same as P1 or P2 (pg 27) – please make clear. • Is 'project ID' here the same as the 'unique project identifier' in the project box?

Use of Outputs by End User

No comments.

Use of Output by Research Community

General Comments:

Significant effort will be required to help the model address the issue of the broader impacts of research (e.g., changes in practice, new policies, and value of the research to the community). If this ultimately falls outside the conceptual model, this gap will need to be explicitly recognised.