Our Data Sources as a Strategic National Asset

Symposium discussion report

Informatics for Social Services and Wellbeing investigators:
Te Rourou Tātaritanga and the Social Wellbeing Agency

9 March, 2023

Introduction

In March 2023, the Informatics for Social Services and Wellbeing Programme, Te Rourou Tātaritanga, and the Social Wellbeing Agency hosted a two-day symposium in Wellington on New Zealand's data assets. Recognising the world-leading status of New Zealand's public data, the symposium brought together national and international data providers, users, experts, and authorities to discuss how the value of the data system can be fully realised and built upon.

The symposium was opened by the Minister for Statistics Hon Deborah Russell. Across two days, world leading international experts in big data infrastructure, analytics, ethics and privacy, presented on a wide range of topics, including keynote speakers Professor Julia Lane (New York University, Cofounder Coleridge Initiative) and Dr Nancy Potok (Director Data Foundation, Former Chief Statistician of the USA). Full speaker biographies and presentation slides are available at `https://terourou.org/symposium2023/speakers/`. Over 150 attendees reflected a broad range of groups involved with New Zealand's data systems, including government agencies, NGOs, data practitioners and academic institutions, alongside the invited international speakers.

Summary

Across the two-day symposium, invited expert guest speakers and discussion participants provided important insights about the effective functioning of administrative data systems, along with aspirations for and issues with New Zealand's administrative data system. There was considerable consistency in the ideas raised across different discussion groups and days. Strong governance of the data system was seen as vital to establishing standards and ethical frameworks, but also drawing together currently fragmented and siloed parts of the system. A lack of understanding and implementation of Māori data sovereignty was considered an issue that

needs to be addressed, along with a lack of trust and social licence among the public. Clear communications about the purpose of the data system and orienting it toward public good research that provides meaningful information to communities was seen as one way to potentially increase trust. Slow processes across the system were seen as a barrier to its efficient use, including being able to inform policy-based decision making in a timely manner. Finally, it was often considered that multiple aspects of the system could be made more accessible and equitable. This could ensure better access across groups and locations, increased input at all levels of the system from community groups, iwi, and NGOs, and better implementation of tools and software to enhance ease of use of the system.

Symposium discussions and key findings

The purpose of this report is to briefly summarise the discussions that took place amongst attendees and participants on each day of the symposium. On Day 1, attendees were divided into three groups based on their area of interest (analytics, policy, or systems) to discuss two key questions: "What functions should an administrative data system have?" and "What are the current issues facing NZ's data system?". On Day 2, workshops took place among five smaller groups of the symposium attendees, who considered the questions: "What issues need addressing in NZ's administrative data system?" and "What are the key components of a national integrated data system?". The remainder of this report presents the key themes in these discussions (based directly on notes taken during the discussions), but not an exhaustive list of all points raised.

Day 1 discussion summary

Day 1 of the symposium was a combination of keynote addresses and presentations by invited speakers and two breakout discussion sessions taking place in-between presentations. These discussion sessions involved both attendees and invited speakers, who collectively contributed their thoughts in large group settings regarding two key questions about New Zealand's data systems. As there were a large number of attendees, discussions took place in three different interest groups: analytics, policy, and systems. The key questions and discussions are summarised below.

What functions should an administrative data system have?

Table 1 provides an overview of the key themes from the Day 1 discussions about functions of an administrative data system. As discussions took place in different interest groups, key functions are attributed to the specific group raising them. Community engagement was a key theme among each of the discussion interest groups. It was considered central to clearly communicate the purpose and use of administrative data, particularly how it can be used to answer local, community relevant questions. This was also considered important for building trust. Strong governance (reflecting clear standards, accountability, and also Te Tiriti principles) featured commonly, as did the need for high quality metadata. While some ideas were raised across multiple interest groups, others were specific to particular groups. For example, it was noted by the Systems interest group that an administrative data system should have a clearly defined purpose of use, while the Policy interest group felt an administrative data system should be oriented toward public good and informing decision making.

Table 1: Key themes from Day 1 symposium discussions about the functions an administrative data system should have.

Key function	Interest group raising point	Description
Ethical	Analytics, Policy	There should be ethical considerations and standards from end to end — from data collection (e.g., informed consent) through to researcher compliance with their original research plan. Knowing where bias is or can be introduced or amplified (e.g., by machine learning).
High quality metadata	Analytics, Policy	Good documentation of metadata, including knowledge of data cleaning that has taken place, who has produced it, exactly what has been measured, inclusions/exclusions. Accesses to data collection forms.
Transparent analytic processes	Analytics	Transparent analytic processes, including acknowledgement of data limitations and documentation of decisions made. May include the reuse of data and code, version control, and open and peer review of code.
Strong governance	Policy, Systems	Governance should be transparent with clear standards in place. Principles of Te Tiriti should be incorporated in governance, and oversight and accountability should be built in (including equal evaluation of system users).
Timely	Policy	Systems should be able to produce required data in a timely fashion and in a way that provides good value.
Accessible	Policy	Training programmes for people using the data and available tools, good support for data literacy. Systems should be as easy to use and understand as possible.

Community engagement	Analytics, Policy, Systems	Purpose and use of data systems should be clearly communicated to the community to build trust. The benefits of using the data, and how it could be used, should be clear. System should be oriented toward answering local questions with good community collaboration, so they can see the value of the data being collected. Incorporating community in process of analysis (e.g., feedback loops).
Public good focused	Policy	The data system should be aimed at supporting the needs of the country and improving the lives of its citizens. Data use should be for creating value and informing decision makers (for funding decisions).
Sustainable and scalable	Policy	The data system should be both sustainable and scalable, and "forward thinking" with adequate funding to support these features.
Defined system purpose	Systems	The data system should have a clear purpose for service, leading to efficient and targeted use (e.g., helping to inform more effective interventions).
Interconnected agencies	Policy, Systems	Many industries and agencies need to address or have similar questions, and could function more effectively in the same ecosystem (e.g., minimise duplication/same data being collected across industries).

What are the current issues facing New Zealand's data system?

Table 2 summarises the key themes from the second Day 1 breakout discussion that took place around issues facing New Zealand's data system. Several issue themes were again evident across each interest group, including a lack of knowledge and implementation of Māori data sovereignty, public trust, and disconnection between different parts of the system. Other key issues commonly discussed included slow and untimely processes in the system, unreliable funding and support, and computational power and software deficits holding the system back.

Table 2: Key themes from Day 1 symposium discussions about the current issues facing New Zealand's data system.

Key issue	Breakout group raising issue	Description
Timely processes	Analytics, Policy	System alignment requirements with how people need to use the data. Output checking and refresh alignment with when the data is needed. Better alignment between refreshes and Ministerial deadlines could be beneficial. There could be routine reporting and readily available summary statistics from the Integrated Data Infrastructure (IDI).
Unreliable funding and support	Analytics, Systems	Funding for projects tends to drop off over time without proper follow through. Available resources vs. technical capability are often mismatched (e.g., improvements that could be made to the systems either don't have the funding or don't have the people to do it).
Lack of knowledge and implementation of Māori data sovereignty	Analytics, Policy, Systems	Te Tiriti is not appropriately realised in the data system. A lack of understanding of what Māori data is in practice (definitions and classifications), or how to implement it.
Lack of trust/social licence	Analytics, Policy, Systems	Public may push back against data systems if appropriate consent isn't in place. This may become increasingly problematic over time. There is a need to monitor and understand public trust and social licence.
Disconnection across system	Analytics, Policy, Systems	There is a separation between data providers and users of data, and uneven governance of data. Users don't always see limitations of data. Different organisations have different boundaries and standards making sharing, pulling together of data, difficult.

IDI literacy	Analytics, Policy	Ability and need to use various software (e.g., SQL) can create barriers to entry (although some noted that lower barriers could lead to lower trust). People have different personal analytic standards when analysing data. Decision makers also often don't have a good understanding of the data.
Gaps in data coverage	Analytics, Policy	Key collections are missing, and data does not capture households well. The system and analytics are individual-based and not whānau/family based.
Computational power/software deficits	Policy, Systems	There could be an API for the IDI. Computational and software restraints are limits to analytical capacity.
Infrastructure does not serve all users equally	Systems	Infrastructure is designed more for statisticians than computer scientists. Structure of data provided not necessarily set up for different uses (machine learning, longitudinal designs). Data is often only collected for the needs of a given agency, but could be made more widely useful.

Day 2 discussion summary

Day 2 of the symposium was similarly a combination of presentations from invited experts and two breakout discussion sessions. Day 2 was more interactive, with shorter presentations and greater emphasis on discussions between workshop participants, which included the invited guest speakers. Unlike Day 1, the two breakout discussions did not take place according to areas of interest, and involved smaller workshop groups, who collated their ideas together before sharing these with the broader workshop group. The following section summarises these discussions.

What are the key components of a national integrated data system?

Table 3 summarises the themes from Day 2 symposium workshop participants' ideas about the components that are key to a national integrated data system, in no particular order. Overall, clearly established governing systems were often cited, with the importance of a coherent system with appropriate established ethical and operational standards emphasised. Having accessibility built into the data system (through the use of tools and collaboration among researchers) was seen as important. Inclusivity was frequently mentioned, such that it was considered important for a wide range of groups outside of government to have input in (and access to) the system. A strong value proposition was also considered vital. Participants felt communication of the uses (and therefore benefits of) New Zealand's data systems were imperative for establishing public trust and social licence. However, they also felt a strong value proposition was necessary for government, demonstrating how the data system can help inform policy and decision making,

underscoring the importance for government to supply high quality data but also appropriate funding.

Table 3: Key themes from Day 2 Symposium workshop discussions
about the key components of a national integrated data system.

Key component	Description
Strong governance	Leadership that can recognise and champion the value of the system, particularly in the policy environment (informing policy needs). Emphasising stewardship, not ownership, of the system. Clear standards/charter for use of linked data.
Transparency	Transparency with processes across the system, especially in communicating the use and value of the system to community and stakeholders. Honesty in limitations of system, such as gaps in collected data.
Accessible	Reusable software and tools available for greater ease of use of data, and more training initiatives. Promoting sharing of code, notes, among researchers, and consistent, detailed metadata.
Inclusive and equitable	System should involve input from various communities (not just government), from iwi, businesses, providers, and public sector groups. Equitable access to data for those who need it.
Centralised and interconnected	A national "system of systems", including standardised systems and processes, including for the sharing of data and information. There should be an ecosystem between providers, users, and communities, with research based on effective partnering between groups.
Ethical	The system should have clear ethical standards ensuring safe, high-quality data that puts people first. Reflects data stewardship and incorporates principles of Māori data sovereignty and Te Tiriti.
Strong value proposition	Value of the system should be clear to communities and individuals, which may help build trust and social licence. Relevance of data and research to policy (and therefore importance of providing quality data) should also be clear to government.

What issues need addressing in New Zealand's administrative data system?

Finally, Table 4 summarises themes in the issues discussion groups thought need to be addressed in the administrative data system. Key issues were identified across a wide range of aspects of the data system, including inequitable access, inadequate data coverage, and siloed systems (involving both data providers and users) which create inefficient use of and processes within the system. The process of extracting meaningful

data and insights from the system was considered too slow and out of sync with government ministry expectations about what data can be provided, and when. A distrust of the system among the public was frequently cited, possibly linked to a lack of understanding of how their data is used. Participants also felt that the Government does not always appreciate the importance of the data and how it can be used for policy decisions.

Table 4: Key themes from Day 2 Symposium workshop discussions on issues that need addressing in New Zealand's administrative data system.

Key issue	Description
Barriers to and inequitable access	The IDI requires a high bar of capability to use and has limited/no external training opportunities, creating a reliance on existing networks already using the IDI. Geographic restraints (availability of local Datalabs) create inequitable access, particularly for iwi/Māori.
Lack of trust/social licence	There is a lack of understanding among the public about how data is used and stored. Perceptions are also changing regarding acceptable use of data (for example, regarding Māori data sovereignty).
Slow and untimely processes	Administrative data is often needed or requested for rapid turnaround, but slow processes across the system (including access, analysis, data availability) hinder this.
Lack of and inconsistent metadata	A lack of information provided with the data makes understanding the data difficult and time intensive. Also a lack of clarity on data quality.
Underappreciated importance of data	Government may not always understand how reliant policy is on the data, while data providers may not fully understand the responsibility of providing quality data. Can create difficulty in making a case for investment.
Disconnected, decentralised systems	Across all aspects of the system, from siloed data providers, researchers (not joining together on projects), and funding.
Gaps and bias in data coverage	Much of the data has biases, regarding the type of data collected and who it is collected from. It does not always capture data that is meaningful to or useful for communities.

Acknowledgements

Thank you to our invited international speakers:

- Professor Julia Lane (New York University, Cofounder Coleridge Initiative)

- Dr Nancy Potok (Director Data Foundation, Former Chief Statistician of the USA)

- Professor Janeen Baxter (Director of the Australian Research Council Centre of Excellence for Children and Families over the Life Course, University of Queensland)

- Professor Mark Western (Professor of Sociology, Founding Director of the Institute for Social Science Research, University of Queensland)

- Bill Simpson-Young (CEO Gradient Institute)

- Professor Chris Dibben (Director Scottish Centre for Administrative Data Research, Co-Director International Population Data Linkage Network)

- Professor Graham Williams (Chief Scientist, Software Innovation Institute Australian National University)

- Professor Peter Christen (College of Engineering, Computing and Cybernetics, Australian National University)

- Professor Tony Hosking (Director of the School of Computing at the Australian National University)

- Glenn Archer (Australian National University, formerly Australian Government Chief Information Officer, head of the Australian Government Information Management Office, and Research Vice President (Government) at Gartner).

Full speaker biographies and presentation slides are available at `https://terourou.org/symposium2023/speakers/`.