



13 November 2018

*This is a submission on behalf of the Artificial Intelligence Forum of New Zealand to Stats NZ on the public discussion document Towards new data and statistics legislation*

Background on the AI Forum NZ:

Te Kāhui Atamai Iahiko o Aotearoa (Artificial Intelligence Forum of New Zealand - AI Forum) is a purpose-driven, not-for-profit, non-governmental organisation (NGO) that is funded by members. The association was founded in 2017.

The AI Forum brings together New Zealand's community of artificial intelligence technology innovators, end users, investor groups, regulators, researchers, educators, entrepreneurs and interested public to work together to find ways to use AI to help enable a prosperous, inclusive and thriving future for our nation.

The Forum advances New Zealand's AI ecosystem through connections, advocacy, growing talent and collaboration. The AI Forum promotes the economic opportunities raised by AI, supporting great applications of AI and emerging New Zealand AI firms, and also works to ensure that society can adapt to the rapid and far-reaching changes that AI technology will bring.

The AI Forum is part of the NZ Tech Alliance. The New Zealand Tech Alliance is a group of independent technology associations from across New Zealand that work together to ensure a strong voice for technology.

CONTACT:

Ben Reid

Executive Director

E: [Ben.Reid@aiforum.org.nz](mailto:Ben.Reid@aiforum.org.nz)

M: +64 27 344 6808

# Questions asked in the public discussion document

## Outcomes for data and statistics legislation

- Government-held data is a strategic asset when used safely and responsibly to improve the lives of all New Zealanders.
- Iwi and Māori rights and interests are actively protected when collecting, managing, and using data.
- Official statistics are relevant, reliable and impartial.
- The independence and integrity of official statistics are actively protected and promoted.
- Government has ongoing access to the data it needs.
- The burden of supplying data is minimised for individuals, organisations, and businesses.
- Data is made open (anyone can freely access, use, and share it), whenever possible, to maximise value and access. When data can't be made open, but can be safely shared through controlled access, it will be.
- Privacy, confidentiality, and transparency underpin collection, management, and use of data.
- Governance and accountability arrangements ensure safe and appropriate use of data to create value, while maintaining New Zealand's trust and confidence.

1. Do you think these proposed outcomes are the right ones for new data and statistics legislation? Please comment on any of these outcomes, and/or list any other outcomes you think should be considered.

The AI Forum generally supports each of these outcomes, with the following comments.

- The principle that data is a “strategic asset” requires clearer definition:
  - while data is important it's actually what we do with it that counts. In that sense, data may be slightly more analogous to a renewable raw material rather than an asset per se. We would suggest more of a holistic approach that emphasises the role of data in an innovation ecosystem that includes skills, applications, investment, etc. to deliver value - rather than identifying data as a standalone “asset”.
- the wording of the first outcome is a bit awkward: “Government-held data is a strategic asset when used safely and responsibly to improve the lives of all New Zealanders.” This carries an implication that, if the lives of only some New Zealanders is improved, then Government-held data is somehow no longer a strategic asset. Consider rewording this outcome.
- Government absolutely requires high quality, accurate and timely data to delivery high quality services and make better evidence-based policy.
  - We would recommend carrying out a "Data Inventory" across publicly owned and publicly stewarded data and publishing and maintaining this catalogue.
  - We support a data architecture for government which supports the “tell

- government once” principle
- Equally important is a data skills inventory across government and support for recruitment and professional development of staff.
- The burden of supplying data should be minimised across the board - and new data sources together with automated technologies (eg Internet of Things, traffic counting systems, satellite imagery etc) should be actively explored to continually reduce the cost and friction involved in collecting data.
- Open Data: in general we recognise a value-multiplier for New Zealand from sharing public data openly and accessibly, understanding that appropriate principle-based sharing constraints may need to be applied. Efforts should be made to accelerate publication of datasets, with priorities driven by market- and social- demand signals. This acceleration of open data publishing will likely require investment.
- Data governance arrangements should include representation from a broad range of stakeholder groups, not just government.

We also note:

- New Zealand needs a joined up, all-of-Government approach to stewarding public data and removal of existing “data silos” unless these are absolutely required to maintain confidentiality, privacy or other data stewardship requirements.
  - This approach would require more than just a new strategy to implement.
  - It may be useful to look towards Estonia for inspiration. They have a system whereby the Govt may only ask citizens for information once (“tell government once” principle), and where citizens can access their data online and see which official has been looking at it.
- Machine Learning and AI techniques generally provide new tools to unlock hidden value within and from combining Data assets - we encourage Stats to work with AI researchers and developers to supply data in usable formats for machine learning applications.
  - We should also consider what is lost when we're unable to access data on a reasonable basis - e.g. it's harder to create better public services and there is likely to be a slower increase in productivity - points made in the recent Algorithm Review. While it's important that there are strong privacy and security protections, if data is held like a reference library that's more interested in holding its books rather than supporting readers, then progress may be limited from what is possible.
- Data may have as-yet-unrealised value in the future - we would advocate that publicly stewarded data could be kept indefinitely unless there are reasons why it should be deleted (for example to comply with Privacy regulation).
  - We cannot know in advance what we might like to do in future, so pre-determining each and every purpose so that we can quality assure our data collection and analysis procedures, while a laudable goal, is not feasible.
  - As a corollary to this, when private data is identified as needing deletion, could the option be explored to enable the owner of that data to preserve it rather than have it destroyed?
  - Data is more like a renewable resource as you can reuse it. It's also worth noting

that only a subset of data is personally identifiable information - see the useful definition of data in this from research firm Sapere

<http://www.srgexpert.com/publications/our-people-publicat-512/>

- Data quality and timeliness are important factors and investment will be needed to be made to improve both.
- Nowhere in the document are data provenance or data audit mentioned. These “metadata” functions are crucial to the wider data functions envisioned for the future. ‘Transparent/transparency’, as defined, does not equate to assuring and explicitly describing provenance. ‘Interpretable’ gets somewhere near it, but is again not quite the whole story. Particularly in order to ensure that open data is used with integrity, detailed provenance info should be provided with all data.
- Related to the last point, if others are to be allowed to use open data provided by Stats NZ then there must be expectations/obligations around their reporting that data and their use, in terms of an audit trail and requiring and supporting clear traceability – otherwise we run risk of someone saying ‘Oh the results are reliable because I got the data from Stats NZ’, even though they may have pre-processed the data beyond recognition.

2. How do you think the Treaty of Waitangi should be recognised across the government data system?

No response

3. How do you think iwi and Māori interests in collecting, managing, and using data should be recognised?

- With respect to the needs and expectations of Māori there may a need to consider specific provisions to enable Māori-only access to certain Māori data.
  - However we would hope that this need would be balanced with the overarching general principle of open data access.
- We would suggest the development of a 'data ecosystem' that draws upon bicultural competence and which reflects genuine bicultural partnership.

## Leading the Official Statistics System

4. Do you agree or disagree with the proposed functions, duties, and powers of the Government Statistician listed above? Please comment.

No response

5. Do you think there are any other functions, duties, or powers for leading and coordinating the Official Statistics System the Government Statistician needs to have?

No response

## Professional independence and ministerial oversight

6. What are your suggestions for ensuring transparency, trust, and integrity in the production of official statistics across government?

No response

## New Zealand's Most Important Statistics

7. Do you think there should be an opportunity for public input when deciding on New Zealand's most important statistics. Please explain.

No response

8. Do you agree that high quality statistics produced outside of government should be able to be recognised as reliable and trustworthy? Please explain.

- In this age of exponentially increasing data quantities and data sources, consider if in future New Zealand only recognised statistics and data produced inside of government then (1) it will miss out on significant value from considering many other non-governmental data sources and (2) will experience significant capacity constraints inside the agencies responsible for collating those official statistics.
- Alternatively, if Stats was instead tasked to provide some level of accreditation as to the quality of 3rd party data sources then that would potentially increase the pool of available data and statistics for both government - and non-government - use and accelerate the value to be gained for New Zealand.

## Survey and administrative data

Administrative data is data collected or created for administrative purposes such as registration, service delivery, transactions, and record-keeping.

Survey data is data collected by directly asking an individual, family, organisation, or business to provide answers to questions.

9. What do you think about the Government Statistician being able to choose the best data source (administrative data or survey data) and require the data to be provided?

No response

10. Do you have any suggestions about what the Government Statistician should consider when deciding the best data source needed to produce official statistics?

No response

## The Census

11. Do you think public consultation should be required before decisions are made on new or altered content for the Census? Please give reasons.

No response

## Open data

Open data is data anyone can freely access, use, and share.

12. What things do you think are important when deciding to make data open?

- We support the desired outcome: Data is made open (anyone can freely access, use, and share it), whenever possible, to maximise value and access. When data can't be made open, but can be safely shared through controlled access, it will be.
- Therefore instead we would start from the position that Data is made open by default and instead the question to be asked is "what reasons should prevent data from being made open?"
- Considerations including national security, individual privacy, likelihood of harm and commercial sensitivity would be valid concerns to prevent making data open, applied on a case by case basis.

## Sharing data for research and analysis

Research and analysis generates new knowledge about: economic, social, and environmental issues; groups of individuals, households, or organisations; and the relationship between different factors and how they change over time.

13. Do you agree or disagree that new data and statistics legislation should clarify that data can be shared across government so that it can be used for research and analysis, with appropriate safeguards and protections? Please give reasons why or why not.

- New Zealand needs an all-of-government approach to stewarding public data which enables more efficient, joined up service delivery, together with sharing data across government for research and analysis.
  - The principle of "tell government once" if applied to joined-up government service delivery would imply that sharing for research and analysis follows as a natural corollary, subject to appropriate safeguards and protections
  - It may be useful to look towards Estonia for inspiration. They have a system whereby the Govt may only ask citizens for information once, and where citizens can access their data online and see which official has been looking at it.
- We would recommend completing a "Data Inventory" across publicly owned and publicly stewarded data and publishing and maintaining this catalogue
  - Equally important is a data skills inventory across government and support for recruitment and professional development of staff.

- The current functional organisation for public data stewardship seems inefficient and siloed - Statistics, Archives, LINZ and other data-stewarding agencies could potentially be combined into a single Data Steward function - eg "Ministry of Data [and Statistics]"
  - Rather than rush into structural change are there incentives to support better collaboration - for example increasing funding to ensure more and better skilled data stewardship staff are available to support other govt agencies.
- From an efficiency and productivity perspective, such a "Ministry of Data" should be given powers to mandate that all public data be placed into its stewardship, subject to appropriate safeguards and protections

14. What protections and safeguards do you think should apply when organisations outside government want to combine their data with government data for research and analysis?

No response

## Access to government-held data for research and analysis

The Statistics Act 1975 requires the Government Statistician to consider the public interest when deciding whether to allow access to Stats NZ-held data for research and analysis purposes

15. Do you agree or disagree that new data and statistics legislation should clarify the public interest test considerations for access to government-held data for research and analysis? Please give reasons for your answer.

- We support to the principle that: Data should be made open (anyone can freely access, use, and share it), whenever possible, to maximise value and access. When data can't be made open, but can be safely shared through controlled access, it should be.
- If this principle is followed then the Public Interest Test considerations would only be applied in situations when data cannot be made open for specific reasons.
- Considerations including national security, individual privacy, likelihood of harm and commercial sensitivity would be valid reasons to prevent making data open.
- For transparency the legislation should clarify the Public Interest Test considerations.

16. Data sensitivity, likelihood of harm, and public expectations are three possible factors to consider when assessing the benefits and risks of research or analysis using government-held data. What other factors do you think should be considered and why?

No response

## De-identification and confidentialisation

De-identification reduces the risk of spontaneous recognition (that is, the likelihood that a person, place or organisation may be identified without any effort). It typically includes, but is not limited to, removing names, day of birth or death, addresses, and unique personal or business identifiers.

Confidentialisation reduces the likelihood that individuals, households, or organisations can be identified by using statistical techniques such as combining two or more groups (aggregation), and changing the number of respondents in a group (eg rounding or suppressing small numbers).

17. Do you agree or disagree with introducing a risk-management approach to confidentiality settings, balancing benefits against the likelihood and potential impact of identification? Please give reasons why or why not.

- Yes, we support a proactive risk-management approach which balances benefits to be realised from publishing data instead of taking an absolutist, overly cautious approach to confidentiality settings. Government should be enabled to take responsible, balanced risk decisions.
- In scenarios where confidentiality settings are breached despite reasonable precautions having been taken, responsible data stewardship should be enabled to approach this like a “bug” in the system to be patched, ensuring that it doesn’t happen again, rather than be subject to large up-front preventative costs and restrictions.

18. Apart from sensitivity of data, what factors do you think should be considered when assessing the potential harm from releasing less-confidentialised data?

No response



## Approving users and providing secure access

Data labs are secure virtual environments sited within secure rooms that approved users need to visit to access data.

19. What do you think are the issues, if any, of allowing access to data by international researchers? How might these be addressed?

We think that overall there are significant economic and social benefit opportunities from enabling access to New Zealand data by international researchers. By doing this we enable a broader international understanding of New Zealand's situation and encourage global innovation which is better-adapted for New Zealand's specific demography and economy.

We acknowledge that there may be scenarios where allowing access to sensitive or confidential New Zealand data for international researchers may disadvantage New Zealand - but these potential disadvantages should be clearly identified in each case.

20. What do you think are the issues, if any, of approving data labs outside of New Zealand? How might these be addressed?

The issues would be usage of the sensitive or confidential New Zealand data in a way which did not conform to New Zealand's applicable laws. (For example breaking the Privacy Act.)

We would suggest that Data labs outside New Zealand which are using sensitive or confidential New Zealand data under licence should be subject to the same legal constraints as New Zealand-domiciled data labs.

Related to the last point, if others are to be allowed to use open data provided by Stats NZ then there must be expectations/obligations around their reporting that data and their use, in terms of an audit trail and requiring and supporting clear traceability – otherwise we run risk of someone saying 'Oh the results are reliable because I got the data from Stats NZ', even though they may have pre-processed the data beyond recognition.

21. What do you think are the issues, if any, of providing data to reputable international organisations for their ongoing use? How might these be addressed?

The same issues would be usage of the sensitive or confidential New Zealand data in a way which did not conform to New Zealand's applicable laws. (For example breaking the Privacy Act.)

We would suggest that Data labs outside New Zealand which are using sensitive or confidential New Zealand data under licence should be subject to the same legal constraints as New Zealand-domiciled data labs.

Related to the last point, if others are to be allowed to use open data provided by Stats NZ then there must be expectations/obligations around their reporting that data and their use, in terms of an audit trail and requiring and supporting clear traceability – otherwise we run risk of someone saying 'Oh the results are reliable because I got the data from Stats NZ', even though they may have pre-processed the data beyond recognition.

## Transparency

22. What information about access to government-held data for research and analysis do you think should be made publicly available? Please give reasons.

No response

23. Are there other aspects of data collection, management, and use that you think government agencies should be more transparent about? Please give reasons.

Government Agencies holding sensitive, private or confidential data should provide regular evidence of their Data Security policies and practices - including the most recent applicable auditors' reports.

Citizens need to be reassured that Government practices to protect private and confidential data are at least equivalent to those of leading international technology businesses.

## Offences and penalties

24. Apart from the two existing broad obligations to provide information to produce official statistics, and to protect confidentiality of information are there any other obligations you think should be able to be enforced?

No response

25. Do you think the two broad types of obligations should be treated with the same level of seriousness? In other words, is failing to provide information as serious as failing to protect confidential information?

No response

## Further suggestions

Do you have any further suggestions on how to improve the government data system for New Zealand?