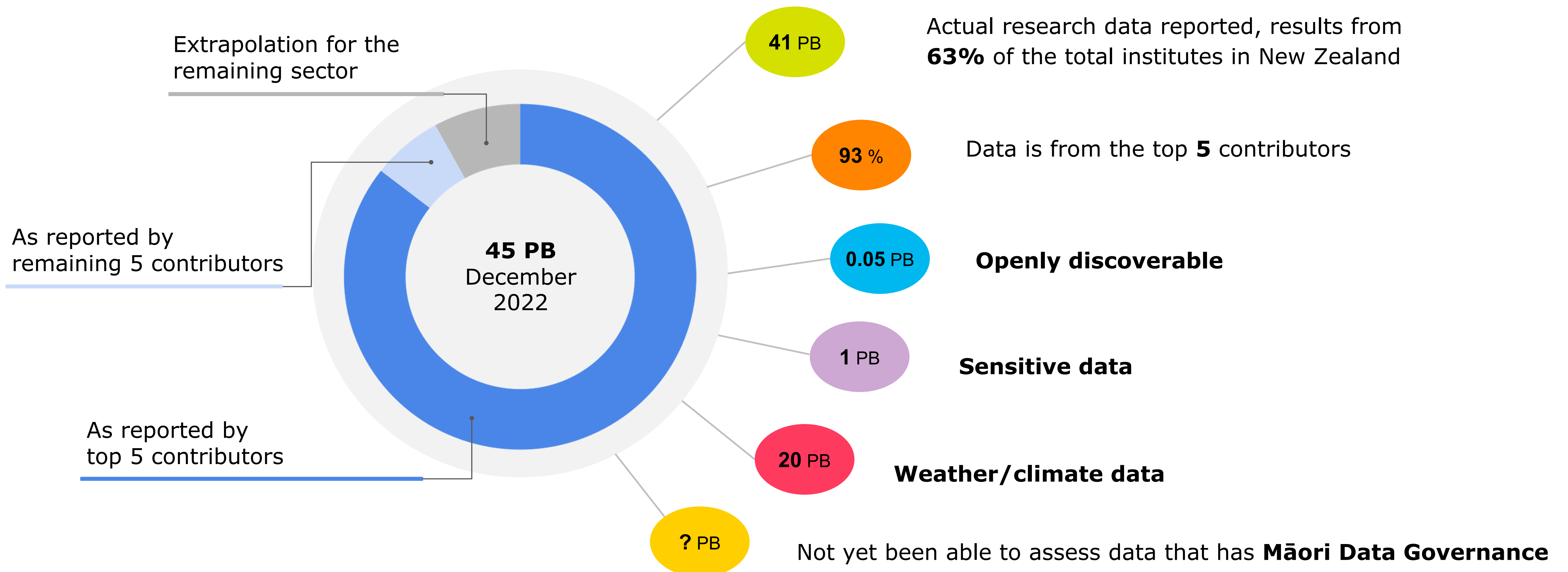


Claire Rye, Nick Jones, Ai-Lin Soo, Luc Betbeder, Rhys Francis



Key Learnings: We don't know much about the research data we keep!

- 45 PB** first copy research data stored for future access by NZ research organisations
- < 1 %** of research data is openly discoverable
- 20 PB weather/climate data** reporting for other disciplines is not possible
- Māori Data Governance** a key aspect of research data in NZ, is not able to be measured

Introducing the research data asset conundrum

Research Data is considered a valuable asset by Researchers and their Institutions alike. Yet very little is actually known about this data which begs further consideration about its value and the culture we have around managing research data.

Working with Research Data Culture Conversation (RDCC) colleagues from Australia, we have been asking a series of apparently simple questions to the Aotearoa New Zealand research data community in order to produce the first 'Macro view' of Research Data as at December 2022.

Starting to measure research data

The RDCC have been exploring questions around Australian research data since 2021, building on earlier conversations sparked by a simple yet fundamental question:

What is the volume of unique data managed by Australian research institutions for future access?

Consulting over 65 research institutions and estimate a research data volume in Australia of 300 PB for Dec 2021.

Translating the RDCC's approach into a New Zealand setting at eResearch NZ 2023 aiming to spark a similar conversation in Aotearoa. We had input from the institutes below, on their respective journeys, and from the wider community.

We have also been seeking to understand the Māori Data Governance of research data. This process needs kōrero with and guidance by Māori to refine how best to engage and ensure Te Tiriti fulfilment within research data held in Aotearoa and to align internationally with initiatives such as Global Indigenous Data Alliance (GIDA).

Estimating Aotearoa's Research Data Holdings

Methodology

Consistent with the RDCC approach in Australia:

- 16 institutions were approached
- 14 agreed to participate
- 12 who have submitted numbers

Participant institutes were asked the following questions:

Question	% Institutes answered
Volume of first copy content held for future access in managed services	83
Volume of data openly discoverable through services provided by the institution	58
Volume of first copy content in services promoted as suitable for sensitive data – meeting ethics and privacy requirements	33
Total storage volume consumed to support the above research data (including replications, etc.)	100

Extrapolating to whole sector is challenging, trialling:

- Research funding
- Publication count
- Citation count
- Field weighted citation impact
- Researcher FTE - best proxy so far
- GDP & R&D expenditure interpolated with Australia

A National Macro View is Possible

We believe this is the first macro view of research data in Aotearoa, giving a whole sector count of 45 PB of research data. The exercise uncovered a few common themes:

- No one could easily answer the questions with a high degree of accuracy
- The "Data person" varied, but always had to consult multiple teams/sources to get answers
- Limited "pain" associated with volume of data
- "Pain" associated with characterisation and management - what is our data culture?
- Even with the refinement in terminology by the RDCC, there was still much confusion over terms, 'sensitive' particularly, was difficult to ensure a common definition was applied.
- Timely as multiple Institutes are currently undertaking their efforts on research data, audits, migrations.

Macro view of research data in Aotearoa

Conclusions from a first Macro View of research data in Aotearoa New Zealand

- Measuring how much research data we have is **difficult**, yet **sparks novel reflections and discussions**
- Differences** between nations are potentially explained by **maturity of managed systems and the scale of research**
- Using Australia's growth rate, **Aotearoa could have ~55 PB** research data in **2023**
- Next steps are being discussed within the research community, with **Māori Data Governance** a key focus

We are very grateful to all the contributing institutes:

