



19 December 2024

9(2)(a) privacy

9(2)(a) privacy

**Official Information Request**

**Our Ref: PSCR 2024-0162**

I refer to your official information request received on 8 December 2024. For ease of reference, I have listed each of your questions and our response directly below.

**1. How does your organisation define Māori data as opposed to non-Māori data**

For data collections at Te Kawa Mataaho Public Service Commission (the Commission), we typically define Māori data, as opposed to non-Māori, using the Stats NZ ethnicity standard.

In the Te Taunaki Public Service Census survey of public servants, we also ask the Stats NZ Māori descent and iwi questions. Collecting iwi information has enabled us to make the data from Te Taunaki more accessible to iwi.

You can find this information publicly available on the Commission's website at the following link: <https://www.publicservice.govt.nz/research-and-data/public-service-census>

**2. Does your organisation have Māori Data Sovereignty and or a Māori Data Governance policy or strategy? If yes, I request a copy.**

**3. I also request a copy of your organisation Data Governance strategy/policy/policies?**

The Commission does not have a separate policy or strategy for Māori data sovereignty or Māori data governance policy or strategy. However, we have incorporated related concepts into our Data Management and Release policy, and its accompanying protocols and guidelines.

Please find enclosed and listed in the table below documents in scope of this part of your request.

Item	Date	Document Description	Decision
1	21 February 2023	Internal Policy – Data Management and Release	Released in full
2	21 February 2023	Data management protocols	Released in full
3	21 February 2023	Data Collection Guidelines	Released in full
4	21 February 2023	Data Release Guidelines	Released in full
5	21 February 2023	Data Release Confidentiality Guidelines	Released in full

**4. Has your organisation had with any success or no/limited success, implementation of any Māori Data Sovereignty Principles or Māori Data Governance? If yes, please provide details of the implementation and how you measured its success.**

For our 2022 review of our Data Management and Release policy we consulted with Stats NZ on how to best incorporate Māori Data Sovereignty principles into our policy. This resulted in the addition of a new core principle in our Data Strategy: *That data use is ethical and culturally appropriate.*

This new principle is explained in the accompanying protocols and guidelines which are listed in the table above. For example, the Data Management Protocols states:

*“Where data is sourced from Māori (or another community such as the disabled or Rainbow communities), or focuses on topics of particular interest to Māori, Māori need to be consulted (e.g., through employee-led networks, population agencies). Ngā Tikanga Paihere: a framework guiding ethical & culturally appropriate data use is a useful resource”*

We have implemented this principle into our process for Te Taunaki, where we worked with the Office for Māori Crown Relations – Te Arawhiti, the Ministry of Māori Development – Te Puni Kōriki, and Te Taura Whiri – Māori Language Commission in the development of the first survey in 2021 and the revisions made for the upcoming 2025 survey. We’ve also engaged with Tūhono (the Māori cross-agency employee led network) in the lead up to the 2025 survey, and are continuing to work with them on the planning for the use of the data.

In addition we also specifically require any researchers requesting access to de-identified subsets of the Te Taunaki data, to outline how their analysis and reporting will support Māori data sovereignty principles in their application.

We have not developed a specific process for measuring the success of implementing these practices.

**5. How many fte are allocated to Māori Data practices in your organisation?**

The Commission does not allocate FTE as per your request. Therefore the Commission has no FTEs allocated to Māori Data practices.

**6. What country/countries are the majority of your organisation’s data stored?**

The majority of the data the Commission’s collects is stored in Australia.

**7. Which Cloud Provider(s) do you use?**

The Commission uses Azure, Snowflake and Tableau Cloud providers.

If you wish to discuss this decision with us, please feel free to contact [Enquiries@publicservice.govt.nz](mailto:Enquiries@publicservice.govt.nz).

You have the right to seek an investigation and review by the Ombudsman of this decision. Information about how to make a complaint is available at [www.ombudsman.parliament.nz](http://www.ombudsman.parliament.nz) or freephone 0800 802 602.

Please note that we intend to publish this letter (with your personal details removed) and enclosed documents on the Te Kawa Mataaho Public Service Commission's website.

Yours sincerely

A handwritten signature in blue ink, appearing to read 'Nicky Dirks', with a stylized, cursive script.

Nicky Dirks

**Manager – Ministerial and Executive Services  
Te Kawa Mataaho Public Service Commission**



## Data Management and Release

<b>Version</b>	2.0	<b>Contact</b>	Strategic Information Team
<b>Policy Owner</b>	Chief Data Officer	<b>Approved</b>	21 February 2023
<b>SharePoint</b>	Strategic Information – Policies & Protocols	<b>Due for Revision</b>	February 2025

### Purpose

This policy provides guidance for Te Kawa Mataaho Public Service Commission kaimahi around managing and releasing strategic data.

This policy and its associated protocols and guidelines are important tools to help the Commission effectively manage our data to ensure the data we release is accurate.

### Scope

This policy applies to all Commission employees (permanent, fixed term and casual), secondees, visitors to the Commission's workplace, contractors and suppliers when at the Commission's workplace or performing business for the Commission. Appendix 1 details data roles and responsibilities.

This policy applies to strategic data only. Strategic data is of strategic value to the Commission in that it either:

- directly contributes to the achievement of the Commission's business strategy and statutory responsibilities
- is high-value public (i.e. non-personal and unclassified) data that the Commission is obliged to release under the [Declaration on Open and Transparent Government](#)
- can be a strategic risk if not managed appropriately
- is useful for agencies across central government
- is useful for multiple business groups in the Commission.

Strategic Data is not:

- internal corporate data (data used to support internal business process or systems)
- documents and records contained in the electronic document management system, SharePoint. These are managed under the Public Records Act 2005.

Strategic data is broadly defined. Te Mana Arataki has decision rights over whether data is defined as strategic data.

## Data Strategy

The Commission's Data Strategy has five principles:

1. *Data is an asset* – Data should be seen as an asset of the Commission, so that it is managed and used to realise its full value to the Commission and our customers.
2. *Data is protected & open* – Data held by the Commission should be proactively shared across the Commission, across the system and with the public unless there are grounds for refusal (such as for personal, confidential, or classified data).
3. *Data into insight* – Data supports the Commission to monitor, design and evaluate the performance of the system, including the Commission itself.
4. *Data is well managed* – Data is managed efficiently and effectively, and can be trusted as being accurate.
5. *Data use is ethical and culturally appropriate* – Our data protocols focus on the well-being of the people, communities, and agencies that supply our data.

The fourth principle underpins the others. For data to be an asset that generates insight and can be proactively released, it needs to be managed well.

## Protocols and guidelines

Strategic data should be managed in accordance with the protocols and guidelines established to support implementation of this policy.

- Protocols are the system of rules and practice which ensure data is well managed.
- Supporting the protocols are guidelines on collection, release and confidentiality which contain a series of steps for users to follow when collecting and releasing data at the Commission.

These protocols and guidelines draw on Stats NZ and digital.govt.nz principles and protocols<sup>1</sup> where appropriate.

## Data Management Protocols

The protocols that guide Commission kaimahi in managing data are that we:

- embed our data roles and responsibilities
- protect the confidentiality, privacy and security of our data
- invest wisely in our data resources
- assure our data quality
- share a data language
- use our data wisely.

The protocols are fully described in this document: [Data Management Protocols](#).

---

<sup>1</sup> In particular, these draw on '[Principles and Protocols for Producers of Tier 1 Statistics](#)', '[Data confidentiality principles and methods report](#)', the [Data Protection and Use Policy](#) and '[Ngā Tikanga Paihere: a framework guiding ethical and culturally appropriate data use](#)'.

## Data Collection Guidelines

The data collection procedures formalise governance around the data collections in the Commission. These are:

- data is collected to meet a specific purpose
- data collections are well-managed, and we invest in strategic data to enhance the value of our data
- new collections are designed well
- appropriate channels are used to collect data.

The guidelines are fully described in this document: [Data Collections Guidelines](#).

## Data Release Guidelines

These procedures formalise governance around the public release of data from the Commission. These are:

- high-value data is published where possible
- published data is secure before release
- published data is well managed.

The protocols are fully described in this document: [Data Release Guidelines](#).

The data release guidelines should be used in conjunction with the Commission's [Official Information Act \(OIA\) Guidance](#).

## Data Confidentiality Guidelines

These guidelines make this possible by allowing as much high-value data as possible to be released, while ensuring that it is:

- not in a form that could reasonably be expected to identify an individual, or
- at a level of aggregation where the data is still informative.

The guidelines are fully described in this document: [Confidentiality Guidelines](#).

## Appendix 1: Data roles and responsibilities

ROLE	DESCRIPTION	KEY RESPONSIBILITIES
SENIOR DATA STEWARD	A role held by the Chief Data Officer, who has oversight over data management and release across the Commission on behalf of the Commissioner.	<ul style="list-style-type: none"> <li>Manages risks around high-value public data releases</li> <li>Reviews data management and release policies</li> <li>Reviews and sponsors new strategic data projects</li> <li>Communicate and promote the value of data</li> </ul>
TE MANA ARATAKI	The Governance group who upholds the principles of the Data Strategy in relation to decisions around managing the Commission's strategic data.	<ul style="list-style-type: none"> <li>Approves the Commission's data management and release policies</li> <li>Makes investment decisions around improving our data resources</li> <li>Approval over what is 'Strategic' data (and therefore subject to all controls)</li> </ul>
DATA OWNER	Has legal rights over the data, including copyright ownership. For the Commission (as a department), all data developed internally or acquired via legislative provision, is owned by the Crown. Data supplied by holders of external copyright remains in their ownership, regardless of the usage licence acquired by the Commission. Ownership of a data system does not equate to ownership of the data.	<ul style="list-style-type: none"> <li>Provides permission for the use and reuse of their data, either through written consent, or the use of creative common licences or similar copyright</li> </ul>
DATA STEWARD / BUSINESS OWNER	Stewardship duties belong to business units within the Commission that are most concerned with a particular data resource. This is a business function, not a technical one.	<ul style="list-style-type: none"> <li>Adheres to the data management and release policies</li> <li>Establishes data requirements and supports procurement of data resources</li> <li>Sets expectations for how data is to be managed (including security, privacy and access settings)</li> <li>Ensures data is used in accordance with the purposes for which it was collected (including adherence to copyright)</li> <li>Approval and contribution to open data releases</li> </ul>
DATA CUSTODIAN	The technical specialists who develop and administer data management practices. The Strategic Information Team is generally the custodian on behalf of the Commission, although external service providers managed via contract (e.g. TMIS) can also be the data custodian.	<ul style="list-style-type: none"> <li>Develops and administers the Data Strategy and data and release policies</li> <li>Advises on, and monitors the use of, the Commission's data management and release policies</li> <li>Sources data and metadata, in terms of content, quality, and conformance with standards and the Public Records Act 2005</li> <li>Ensures that appropriate Service Level Agreements are in place with platform and data services support providers</li> <li>Ensures expectations and rules set by the Data Steward are appropriately implemented</li> <li>Coordinates release of data to agencies to drive performance improvement and the public under open data requirements</li> <li>Supports data creation through day-to-day service provision and maintenance of data systems</li> <li>Support Te Mana Arataki in its role, in particular in assessment of investment in strategic data</li> </ul>
COMMISSION DATA USER	The Commission staff who analyse data, as part of their day-to-day work, to generate actionable insights.	<ul style="list-style-type: none"> <li>Knows their responsibilities with data as set out in the Commission's data protocols</li> <li>Must adhere to the privacy and security rules and guidance set out in the Commission's data protocols, in particular when externally releasing data</li> <li>Only has access to the data that they need for their role</li> <li>Must adhere to copyright obligations</li> </ul>
EXTERNAL DATA USER	Those looking to the Commission data to provide them insights into the public sector.	<ul style="list-style-type: none"> <li>Must adhere to Crown copyright obligations</li> </ul>

## Data management protocols

PROTOCOLS	PRACTICES
1. WE EMBED OUR DATA ROLES AND RESPONSIBILITIES	<ul style="list-style-type: none"> <li>• Data is an important asset with appropriate governance structures</li> <li>• Data sources each have a steward responsible for life cycle management</li> <li>• All staff know their responsibilities with data</li> <li>• These responsibilities also apply to external researchers given access to the Commission data</li> </ul>
2. WE SECURE THE CONFIDENTIALITY AND PRIVACY OF OUR DATA	<ul style="list-style-type: none"> <li>• Legal and ethical obligations around managing data are adhered to</li> <li>• The permissions we have to use data are understood</li> <li>• Security practices are built into our processes and infrastructure</li> <li>• Privacy or security breaches are managed openly and quickly, recognising the seriousness of maintaining confidentiality</li> </ul>
3. WE INVEST WISELY IN OUR DATA RESOURCE	<ul style="list-style-type: none"> <li>• All data that is collected has a clear use and its value is understood</li> <li>• Data is actively used, and then archived, to get full value from it</li> <li>• Data is appropriately and efficiently sourced</li> <li>• Data infrastructure is invested in to enhance the value of our data</li> </ul>
4. WE ASSURE OUR DATA QUALITY	<ul style="list-style-type: none"> <li>• There is a culture of professionalism and good practice</li> <li>• Data meets the needs of users, within available resources</li> <li>• Data is accurate</li> <li>• Data is timely enough to be of value to users</li> <li>• Data is consistent</li> <li>• Data is ethical and culturally appropriate</li> <li>• Methods used to produce data are understood and documented</li> </ul>
5. WE SHARE A DATA LANGUAGE	<ul style="list-style-type: none"> <li>• Common data standards are used to manage data</li> <li>• The Commission uses national and international data standards where possible</li> <li>• The Commission promotes common data standards across the system</li> </ul>
6. WE USE OUR DATA WISELY	<ul style="list-style-type: none"> <li>• Data infrastructure assists with turning data into insight</li> <li>• Processes get the right data to the right people at the right time</li> <li>• Published data is presented clearly and supported by analysis</li> <li>• Published data is open and accessible</li> <li>• Published data is understandable</li> <li>• Significant errors in published data are corrected quickly</li> </ul>



## Data management protocols

These protocols are the system of rules and practice that ensure data is well-managed at Te Kawa Mataaho Public Service Commission (the Commission). They should be used in conjunction with the data management and release policy and its associated protocols and guidelines.

### **PROTOCOL 1: We embed our data roles and responsibilities<sup>1</sup>**

#### **1.1 Our data is an important asset with appropriate governance structures**

The Commission's leadership team, Te Mana Arataki, have governing authority over the Commission's strategic data. Key governance responsibilities include making key investment decisions and establishing the data management practices. Te Mana Arataki are supported by the Strategic Information team (SIT).

#### **1.2 Data sources each have a steward responsible for life-cycle management**

All data coming into the Commission has a business owner, called the Data Steward. The Data Steward is ultimately responsible that the data is well-managed.

The Data Steward is supported by Data Custodians. These are technical data specialists who develop and administer data management practices. Data Custodians will generally be SIT but may be the Commission IT or contracted out to an external provider.

All data has a data owner that ultimately has legal rights over the data. Data where the Commission is the data owner, either data that is developed internally or acquired via legislative provision, is owned by the Crown.

#### **1.3 All staff know their responsibilities with our data**

These protocols set out the roles, responsibilities and rules for the Commission staff to adhere to. All staff involved in the use of the Commission strategic data are aware of their obligation to protect confidentiality and privacy; and are aware of the penalties for wrongful disclosure. Staff need to agree and sign the Access and Security Protocols before being given access to identifiable unit record data.

#### **1.4 These responsibilities also apply to external researchers given access to the Commission data**

Bona fide researchers may be given onsite access to de-identified unit record data for research purposes. A condition of granting access is that researchers agree to follow the Commission's data management and release policy as if they were a Commission employee. They need to sign the Access and Security Protocols before being given access to unit record data.

---

<sup>1</sup> Data roles and responsibilities are detailed in Appendix 1 of the Data Management and Release policy.

## **PROTOCOL 2: We secure the confidentiality and privacy of our data**

### **2.1 Legal and ethical obligations around managing data are adhered to**

The legislative and ethical obligations governing the collection, storage and release of data are built into the Commission policies and data practices (the Commission's Privacy Officer is responsible for ensuring the Commission policies and practices conform to legislative obligations).

All staff involved in the use of the Commission's strategic data are aware of these obligations to protect confidentiality and privacy and the penalties for wrongful disclosure and adhere to these obligations.

### **2.2 The permissions we have to use data are understood**

Survey respondents and data owners provide permissions for the use of their data. The Commission only uses data in a way consistent with these permissions and the reason(s) for which the data is collected. Staff need to confirm that they understand and agree to these permissions prior to being granted access to some types of strategic data.

We are sure of the need for the data before asking for it. Respondents are informed of their rights and obligations in providing data. Respondents are clear why we are collecting the data and are aware of our security controls to maintain their privacy and confidentiality. The respondent's confidentiality is always strictly preserved unless they have explicitly agreed to the contrary. We support people to know what data we hold on them and their right to access it.

### **2.3 Security Practices are built into our processes and infrastructure**

Security practices are built into our infrastructure. SIT and the Commission IT administer security practices across our data environment. Identifying data (such as name) should be removed from datasets. Application of the Data Confidentiality Guidelines ensure identifying data is suppressed before aggregated results are released outside the Commission. SIT can help with checking the implementation of these rules.

Security practices explicitly set out the rules for access to private and confidential data by a user. It is the data steward, who sets and documents those rules. The data steward also ensures users acknowledge the rules that they can operate within, including their responsibilities to maintain privacy and confidentiality, and also keeps an audit trail.

### **2.4 Privacy or security breaches are managed openly and quickly, recognising the seriousness of maintaining confidentiality**

As per our Privacy Policy, the Privacy Officer is notified immediately of any potential privacy breaches.

## **PROTOCOL 3: We invest wisely in our data resources<sup>2</sup>**

### **3.1 Strategic data is invested in to enhance the value of our data**

When assessing whether to invest in a new data collection, we must consider if it meets the definition of strategic data. If it meets the criteria, the Commission team must talk to SIT who can help assess whether to invest, and the level of maturity needed, based on several technical considerations that include value for money.

### **3.2 All data that is collected has a clear use and its value is understood**

The value of the data we collect is understood. Data collection has clear objectives and data needs. We balance the need to collect data to inform decision making against the costs of production and the burden placed on the system. Work programmes are periodically reviewed to ensure their relevance and justify their continuation.

### **3.3 Data is actively used, and then archived, to get full value from it**

We maximise the use and value of existing data by integrating or aligning it with administrative sources. The Commission understands (through the Data Architecture) and has shared access to all our data resources to be able to get best value from the data. We share our data with the system and public so that they can derive value from it<sup>3</sup>.

Business processes adhere to disposal and retention policies and rules set out in the Public Records Act 2005. In practice however, most data are retained indefinitely due to their potential research value, subject to security, confidentiality and statutory obligations.

### **3.4 Data is appropriately and efficiently sourced**

We ask for data only once and then share internally. Similarly, Data requests to the same respondents are co-ordinated, in particular to agencies and Chief Executives. Existing data sources are used wherever possible and existing collection mechanisms are considered when looking to capture new data. Appropriate opportunities to reduce costs are actively sought. These include economies of scale, data integration, and methodologies and systems that use generic and/or automated processes.

Data collection is designed in a supplier friendly way and with sufficient flexibility to accommodate changes in user needs. SIT can help develop and build standards for data collection, including the design of survey questionnaires.

### **3.5 Data infrastructure is invested in to enhance the value of our data**

The Commission's existing data infrastructure, referred to as the BI environment, is the enterprise solution for the Commission. Where possible the Commission uses this BI environment to enable better integration of data. When investing in new data collection, storage, or administrative systems the Commission assesses it for compatibility with the existing BI environment. New technologies are routinely investigated to see if they would provide value to the BI environment.

---

<sup>2</sup> Also refer to the Data Collection Guidelines.

<sup>3</sup> Also refer to the Data Release Guidelines.

## **PROTOCOL 4: We assure our data quality**

### **4.1 There is a culture of professionalism and good practice**

Analysts act with integrity, objectivity and comply with the Commission's data protocols. The Commission uses good data and project management production processes. The Commission invests in training and development to ensure analysts have the skills required. The Commission regularly assesses data processes and tools; seeking opportunities to implement new technologies and maintaining a culture of peer review.

### **4.2 Data meets the needs of users, within available resources**

The Commission designs new data to be relevant to users' needs, within available financial resources. Teams understand who the key users of their existing data are and why they use it. They consult their key users before making substantial changes to how the data is collected, managed, or reported.

### **4.3 Data is accurate**

The Commission produces its data using sound data and methods. It does this by validating incoming data, and understanding or, where possible, controlling the level of error in its data. Reports that will be published or ministerial briefings must have any quantitative data checked by SIT, or someone with the requisite knowledge.

### **4.4 Data is timely enough to be of value to users**

To be relevant, data needs to be released in sufficient time to meet key users' needs. Timeliness is a decision involving trade-offs between quality and cost. Regular publicly released data should have planned release dates.

### **4.5 Data is consistent**

Key data results should be reported consistently (i.e., there is one version of the truth). Analysts should check to see whether data has already been reported, and if so, use the same figure. Where data is not comparable, it should be flagged as such to key users. Key users are advised of substantial changes to methods that affect consistency with previously reported data. Consistency will be facilitated through the data infrastructure (e.g., through Tableau).

### **4.6 Data is ethical and culturally appropriate**

Where data is sourced from Māori (or another community such as the disabled or Rainbow communities), or focuses on topics of particular interest to Māori, Māori need to be consulted (e.g., through employee-led networks, population agencies). [Ngā Tikanga Paihere: a framework guiding ethical & culturally appropriate data use](#) is a useful resource.

### **4.7 Methods used to produce data are understood and documented**

Understanding how data is produced sheds light on data quality. Both the methods and classifications used in producing the data and measures of the accuracy of the data (e.g., response rates) should be documented. Documentation is regularly reviewed and updated. Its level of detail suits the needs of its intended audience. Documentation should be standardised across the Commission where possible.

## **PROTOCOL 5: We share a data language**

### **5.1 Common data standards are used to manage data**

The Commission works to apply common frameworks, classification, derivations, and metadata to ensure that data has the same meaning across the organisation and that separate datasets can be related to each other.

- Frameworks are a logical structure for organising complex data (e.g., Treasury’s Living Standards Framework)
- Classifications organise data by grouping similar items into understandable categories (e.g., the Australian and New Zealand Standard Classification of Occupations ANZSCO)
- Derivations are standardised ways to calculate one variable from another (e.g., the gender pay gap)
- Metadata is data that helps provide context around data (e.g., response rates, the proportion of respondents who didn’t answer the question)

At the Commission, our core framework looks at the varying dimensions of performance which can be aggregated and disaggregated between CE, agency, sector, and system.

SIT will be data custodians for data frameworks, standards, classifications. SIT will ensure that this common data language is embedded into the Commission’s practice through Tableau and documented into the Data Architecture. These will be incorporated in all new survey and administrative data collections. They should be incorporated into existing data collections during major revisions or upgrades to minimise costs.

### **5.2 The Commission uses national and international data standards where possible**

To facilitate meaningful comparisons between agencies, sectors, and countries, the Commission’s data frameworks, classifications and derivations should align wherever possible with existing system, national and international standards.

SIT can advise on whether there are relevant existing data frameworks, classifications, and derivations.

### **5.3 The Commission promotes common data standards across the system**

The Commission works cooperatively with other agencies in the development of common data frameworks, standards, classifications, and derivations (e.g., the Workforce Information Standard).

The Commission documents its common data frameworks, standards, classifications and derivations and makes them available to other agencies and supports other agencies to use them.

## **PROTOCOL 6: We use our data wisely**

### **6.1 Data infrastructure assists with turning data into insight**

The Commission has a package of state-of-the-art tools (Tableau, R) that can turn data into insight. The SIT team can advise on using these tools.

### **6.2 Processes get the right data to the right people at the right time**

Data supplied meets user needs, in terms of both substance and timeliness. Published data is secure before it is released (refer to the Release Guidelines). Data requests should come directly from users or should be put in writing.

### **6.3 Published data is presented clearly and supported by analysis**

The Commission data is objective. Conclusions are supported by analysis.

### **6.4 Published data is open and accessible**

The Commission data meets its Open Government obligations<sup>4</sup>. Anyone can freely access, use, modify, and share the Commission published data, detailed data is released onto [open.govt.nz](http://open.govt.nz). Published data is made available to all at the same time, where possible.

### **6.5 Published data is understandable**

Data is presented clearly and simply, with easily understandable conclusions. The decisions behind the type of commentary and analysis provided should be reasonable. Judgement is used to tailor the presentation of the Commission published data to our targeted users. Data that is published for transparency purposes communicates key results in straightforward terms for the wider public.

As much detail as is reliable and practicable is made available, subject to confidentiality constraints. To encourage comparative analysis and to provide context, dissemination should include background data on metadata, trends and links to related data. Recurring data releases are delivered in consistent formats. When the Commission reuses data collected by others, credit is given to the original data source.

### **6.6 Significant errors in published data are corrected quickly**

All published errors are handled in a standard manner consistent with their significance and revised when necessary. Revisions practice is covered in the Release Guidelines.

---

<sup>4</sup> See the [Declaration on Open and Transparent Government](#).

# Data Collection Guidelines

These guidelines set out our practice around data collections at Te Kawa Mataaho Public Service Commission (the Commission). They should be used in conjunction with the Data Management and Release Policy and its associated protocols and guidelines.<sup>1</sup>

## 1. Roles and responsibilities are adhered to

- All staff have a duty to take reasonable care of sensitive data provided to them, this can be physically, electronically or even verbally.
- the Commission staff should know their responsibilities with Commission data and their obligations to protect confidentiality and privacy.<sup>2</sup>
- Prior to being granted access to sensitive strategic data, staff need to acknowledge the rules set by the data steward.

## 2. Data is collected to meet a specific purpose

- The purpose for new data should be understood, with clear objectives before any data collection is implemented.
- The value of the data the Commission collects is understood and weighed against the cost of collection.
- Only data that is reasonably necessary for the specified purpose is collected.
- Data is used as intended, in a way that aligns with the permissions the data owner has provided (in accordance with the Commission's Privacy Policy and the data privacy principles of the Privacy Act 2020).
- Personal, confidential and classified data is protected and respondents understand their rights and obligations in providing data.

## 3. Our data is well managed

- New strategic data<sup>3</sup> collections require Te Mana Arataki approval. SIT can help input into the decision paper for Te Mana Arataki.
- All strategic data needs to be managed in accordance with Data Management & Release policy and its related protocols and guidance.
- The Strategic Information team (SIT) will consider incorporating a new data collection into the managed data environment if:
  - The data is 'strategic Data'
  - The collection will be repeated regularly

---

<sup>1</sup> The [Data Protection and Use Policy](#) at Digital.Govt.NZ is a useful reference when designing a new data collection.

<sup>2</sup> As outlined in Appendix 1 of the Information Management and Release policy.

<sup>3</sup> Strategic information is defined in the Information Management and Release policy.

- The data could be used to link with other data sources
- SIT are unlikely to incorporate a collection into their environment if:
  - The data is 'strategic data' and the collection is an ad-hoc one off
  - The nature of the data is not strategic (e.g., corporate data)
- If the data is not strategic, SIT can still provide advice and support to ensure data is well managed.
- When investing in a new collection, storage or administrative systems, the Commission and SIT need to ensure it is compatible with the current environment.
- When data needs have been defined and the collection approved, a data custodian from SIT will work closely with the business owner (or data steward) throughout the project.

#### 4. **New collections are well designed**

- The Commission teams must work with SIT if the need for new data has been identified.
- The Commission need to manage data collections well. Designing data collections is the first steps in our process and it is important the Commission teams talk to SIT at this stage so SIT can help design and develop tools that are fit for purpose and appropriate data protocols and standards are being used.
- Where collections will source data from Māori (or another community such as the disabled or Rainbow communities), or will ask data on topics of particular interest to Māori, the Commission should engage with Māori before collecting the data (e.g., through employee-led networks, population agencies). [Ngā Tikanga Paihere: a framework guiding ethical and culturally appropriate data use](#) is a useful resource.

#### 5. **Data collections are implemented in a respondent-friendly manner**

- The Commission teams must talk to SIT before asking for data from agencies or individuals as the data may already be captured (and we should never ask for the same data more than once).
- If SIT do not have the required data, we need to consider:
  - The individual providing the data and how much work is involved for them in providing it.
  - The best tool to capture the data in a respondent-friendly manner (e.g. Microsoft Forms).
  - How the data will be used and the questions it will answer, to ensure the right data is captured.
  - Respondent-friendly language is used and statistical jargon is avoided – SIT can provide advice on questionnaire wording.
  - Respondent's are aware of why their information is needed and how it will be used.
  - Any data standards we use – SIT can provide advice on current standards, classifications and derivations.



- How frequently we should ask for the data (the usefulness of timely information should be weighed with the cost to respondents and the Commission of collecting the data).

## 6. **Appropriate collection channels are used**

- The Commission has a number of different collection channels with differing levels of maturity. These are detailed in Appendix 1.
- Identifying the right channel to collect data based on factors such as:
  - Repetitive – is the data collection going to be repeated? If yes, more mature methods are desirable.
  - Content likeliness to change – will content potentially substantially change? If yes, more mature methods provide less value.
  - Analysis complexity – is analysis time consuming and / or difficult to understand? If yes, more mature methods are desirable.
  - Is high quality data essential? If yes, more mature methods are desirable.
  - Security – is data sensitive? If yes, more mature methods are desirable.
  - Turnaround time – Is speed of collection and analysis important? If yes and collection is one-off, then less mature methods will be quicker. However, if the collection will be repeated then more mature methods will result in quicker turn around on an ongoing basis (although initial set-up will take longer).

## 7. **Collected data is validated**

- New tools for collecting data should be tested thoroughly before implementation to ensure they can work within our managed environment and that any bugs are found and fixed. New collections should be tested before being run for the first time, even if they use existing tools.
- When data is collected the data needs to be validated to ensure it has been collected as intended and is consistent with other instances of the collection and any erroneous data is amended in a consistent way.
- To validate the data:
  - Build in validation rules.
  - Check the data has been collected correctly (the correct template and data types).
  - Check for outliers in the first instance – Any specific values that do not look correct should be checked with the data provider, if the value is incorrect, the data needs to be updated.
  - Additional checks should be completed as the data is analysed such as:
    - Large changes over time.
    - Large differences compared with other respondents.

8. **Collections are reviewed regularly:**

- Regular collections are reviewed periodically to establish whether:
  - the data is still needed
  - other appropriate data sources have become available
  - we can improve the method of collection
  - the views of respondents and users of data are being taken into account
  - variables are used effectively and that only necessary data is collected

## Appendix 1. The Collection Maturity Model and our Collection Channels

### Collection Maturity Model

Less Mature	More Mature	Most Mature
Manual input and checking of Raw Data		Automated data collection & validation
Manual integration with other data sources for analysis		Data is integrated with other data sources on collection
Raw data exists in multiple places, no single source of truth	Central source of truth, protections against data loss	Data stored in database
Calculations hard to trace back and repeat	Clear transparent calculation processes	Automated calculations
Data collection is owned by the business unit	Data collection is centralised	Data collection is automated, minimal maintenance required

#### Less mature channels:

- Email – information is typed directly into the body of an mail
- If the data is sensitive, additional security measures should be put in place such as password-protected files; using SEEMAIL with security classification to email .
  - Electronically – Email attachment.
  - Physical Transfer – USB drives, Iron Keys, CDs

#### More mature channels:

- Secure portals that require data submission:
  - Microsoft Forms - Surveys are created in Microsoft Forms and the data captured from respondents is securely stored on the Microsoft Forms server. The data is owned and administered by the survey creator, however Microsoft do have the right to access the survey data to provide the survey creator and respondents support. Microsoft also have the right to share the data you capture in some cases (such as if they obtain your consent or if they aggregate or de-identify the data).  
  
Microsoft Forms allows the survey creator to control who completes the survey. Microsoft Forms is a good option for collecting data from a number of individuals based on specific questions identified to answer a data need. This option would be preferred to collecting the data electronically or physically from each individual and collating the responses together which would be very time consuming and open to error.
  - MS Group – An MS group can be set-up to share files securely between agencies. Access can be restricted, with users needing to go through MS Authenticator for access.
  - Next Cloud - Next cloud can be used for sharing data by loading the data into next cloud and setting up access for users you want to share the data with. The users' username and password should be sent via email independently of the link to the next cloud file.

- CFISnet - CFISnet is used to collect confidential data in a secure manner. The data is uploaded into a CASS managed database once the sender has logged in to RealMe to verify their identity for security purposes. This option has been set up for regular confidential data collection.

CFISnet is used to collect the Workforce data and is the right tool to collect this data because it is a secure file transfer process for large sensitive files.

**Most mature channel:**

- Accessible Systems
  - These are systems that don't require manual steps by users to submit data e.g. APIs that developers can access programmatically.
  - Examples of this include:
    - the SAP API for Talent Exchange data,
    - the DIA Govt. A-Z directory API for agency and ministerial contact data,
    - SFTP direct cross-agency database connections that are being built for the Workforce Data pipeline.

Channels such as these would only be used for strategic data and managed by SIT.

# Data Release Guidelines

These guidelines formalise governance around the public release of data from Te Kawa Mataaho Public Service Commission (the Commission). They should be used in conjunction with the data management and release policy and its associated protocols and guidelines.

## 1. High value data is published where possible

- The Commission meets its obligation under the Declaration on Open and Transparent Government to actively release high value public information.<sup>1</sup>
  - Public demand, such as OIA requests, can help the Commission identify its high value public data.
- To support this obligation, the Commission follows the [New Zealand Data and Information Management Principles](#):
  - Data is open, unless restricted by the Official Information Act (OIA) or other policy.
  - The withholding provisions of the OIA are useful guidelines when considering whether or not to release data (see the Commission's [OIA Guidelines](#)):
    - protect New Zealand's security
    - protect the privacy of natural persons
    - protect information where making it available would be likely to unreasonably prejudice a person's commercial position
    - protect information that is subject to an obligation of confidence
    - avoid prejudice to health and safety
    - maintain the constitutional conventions that protect the confidentiality of advice tendered by Ministers of the Crown and officials
    - maintain the effective conduct of public affairs through the free and frank expression of opinions
    - maintain legal professional privilege, or
    - enable a Minister or department to carry on negotiations without prejudice or disadvantage
  - Data is released proactively and readily available online, either on the Commission's own website or other relevant sites (such as data.govt.nz)
  - Data is free of charge where possible
  - Data is released in a form that makes it easily reusable (i.e. most detail possible, machine readable, with metadata) and accessible
- To support regular publicly released data, planned release dates should be scheduled and published.

---

<sup>1</sup> High value public information is publicly funded data, which when re-used contributes to economic, social, cultural or environmental growth, illustrates government's performance, and contributes to greater government efficiencies through improved information sharing. Public data is non-personal and unrestricted data.

- The Commission shares high value system data and insights with agencies to support system performance.

## 2. **Published data is well managed**

- Published data has been managed under the Commission’s Data Management and Release Policy and Protocols, to ensure that:
  - Published data is trusted and authoritative (i.e. accurate, relevant, timely, consistent and without bias)
  - Personal, confidential and classified data has been protected.
  - The commission develops and shares published data and insights in an inclusive, ethical and culturally appropriate way. For example, where data focuses on Māori (or another community such as the disabled or Rainbow communities), or on topics of particular interest to Māori, the Commission should engage with Māori before publishing (e.g., through an employee-led network or relevant population agency).
    - [Ngā Tikanga Paihere: a framework guiding ethical and culturally appropriate data use](#) is a useful resource.
- Data is checked before release to ensure its accuracy.
  - Reports that will be published or ministerial briefings need to have any data checked by someone with the requisite knowledge, and need to be discussed with the Strategic Information Team, if they substantially rely on data.

## 3. **Published data is secure before release**

- Access is restricted to specified individuals or agencies prior to public release. These are:
  - Those directly involved in production
  - Ministers
  - System leaders
  - Select Committees
  - Data provided early for agency Annual Report publication timelines
  - Agencies who provide the raw data can access their own data along with any system benchmark data
- All others enjoy equal access to published data. It is made available to all at the same time.

## 4. **Revisions are managed openly and transparently**

- Published errors are handled in a standard manner consistent with their significance and revised when necessary. All errors in the Commission’s published data need to be reported to the tier three manager responsible for that data. The tier three manager should classify the error as:

- **An error that significantly alters the meaning of the previously published data.** The tier two manager needs to be informed. This needs to be revised on the website as soon as possible. Users need to be identified as soon as possible.
- **Errors that do not significantly alter the meaning of the published data.** These may be revised with the next regular data release, or if in an irregular release, when it is convenient to the Commission to do so.
- **Errors that do not alter the meaning of published data and that will have very little or no impact on users.** These errors can be ignored (revisions which are frequent and trivial will undermine user confidence).
- For each major data release, we publish and maintain a general statement describing the Commission's revision practices.
- Users should know when planned revisions are due and be aware of them as they arise. Scheduled revisions are managed systematically, pre-announced and are reflected in communication plans.
- All revisions should be accompanied by documentation which explains their nature, provides good analysis of the differences between the original and revised data and explains the effect on any previously published commentary or interpretation.



## Data Release Confidentiality Guidelines

The Data Release Guidelines state that Te Kawa Mataaho Public Service Commission (the Commission) should **release high value** data where possible, while also protecting personal data. These guidelines make this possible by allowing as much high value data available for release, while ensuring that it is:

- not in a form that could reasonably be expected to identify an individual, or
- at a level of aggregation where the data is still informative.

The guidelines are consistent with Privacy Act 2020<sup>1</sup> requirements and were developed in consultation with Stats NZ.

The guidelines apply to:

- all strategic data
- any statistical data that contains private or confidential data.

### 1. Roles and responsibilities are adhered to

- All Commission staff have a duty to take reasonable care of sensitive data provided to them, this can be physically, electronically or even verbally.
- The Commission staff should know their responsibilities with the Commission data and their obligations to protect confidentiality and privacy by reading the data management and release policy and its associated protocols and guidelines.
- Prior to being granted access to sensitive data, staff need to acknowledge the rules set by the data steward.

### 2. Use of confidential data within the Commission

- Even if data is not going to be disclosed publicly, or external to the Commission, privacy and confidentiality should be maintained.
- Data should be provided to users at the highest level of aggregation, and/or anonymised, while still being informative for the purposes of the user.
- When providing data to users that is private or confidential in nature, the user should be told this, and if necessary ensure that they know their roles and responsibilities.

### 3. Data release rules

- All Commission staff should follow the release rules provided in Tables 1 and 2 in the Appendix. The release rules provide statistical advice on aggregation, suppression and counts.
- Data release rules apply to all personal data. All Te Taunaki | Public Service Census data is considered personal data. For Public Service Workforce Data from agency administrative data, personal data is defined in Table 3 in the Appendix.
- The Commission can choose, in the following circumstances, not to apply the rules:

---

<sup>1</sup> Privacy Act 2020 is available on [www.legislation.govt.nz/act/public/2020/0031/latest/LMS23223.html](http://www.legislation.govt.nz/act/public/2020/0031/latest/LMS23223.html)



- When supplying data back to the original supplier of the data. For example, supplying agency workforce data back to the Human Resource team of that agency.
  - If data is already publicly available or can be easily ascertained by a member of the public. For example, the gender of senior leadership team members.
  - When reporting data that is organisational, rather than personal, in nature. For example, counts at an agency by job title or occupation level. See Table 3 for examples of personal data.
- **Note** Stats NZ advises that gender pay gaps are not statistically robust for groups of fewer than 20 men and 20 women. Gender pay gaps for smaller groups can be released, if they meet the other rules in these guidelines and the release is accompanied by reference to this data not being statistically robust according to Stats NZ.
  - Note that the Commission has an agreement with the Government Communications Security Bureau (GCSB) and New Zealand Security Intelligence Service (NZSIS) to not release their data collected in the Public Service Workforce Data, beyond that published in these [three tables](#), for national security reasons.

## Appendix:

**Table 1 – the Commission Data Release Rules**

	<b>Counts (e.g. Headcount, FTEs)</b>	<b>Magnitudes (e.g. Salary, Sick Leave)</b>	<b>Likerts*</b>
Full coverage (e.g. Workforce Data and Te Taunaki  Public Service Census)	<p>Suppress cells with counts below '4' (this takes the possibility of collusion into account). Make them appear like cells with zero counts (i.e. either with a zero or by leaving blank).</p> <p>Totals and percentages should be calculated using suppressed and rounded numbers.</p> <p><b>Additional rule if needed</b> Apply random rounding to base 3 by using this <a href="#">workbook</a> if suppressed values can be determined by looking across tables or a highly sensitive variable is involved (e.g. income). Note any use of random rounding.</p>	<p>Suppress cells with counts below '4' (this takes the possibility of collusion into account). Make them appear like cells with zero counts (i.e. either with a zero or by leaving blank).</p> <ul style="list-style-type: none"> <li>• Round magnitude values to an appropriate level. For example, round average salary to the nearest \$100.</li> <li>• Round average tenure, age and sick leave to one decimal place.</li> <li>• Percentages should be calculated using rounded numbers.</li> </ul> <p>For Medians &amp; Percentiles apply Table 2.</p> <p><b>Additional rule if needed</b> Suppress cells with counts below '10' in the following circumstances: The use of multiple variables (e.g. occupation by department) or detailed variables (e.g. occupation at very detailed levels) means the risk of identifying an individual's data is high and the value is highly sensitive (e.g. income).</p>	<p>Results for questions with counts of less than 10 are suppressed with a 's'. This rule is applied at the question (denominator) level, not the response (numerator) level.</p> <p>This is the approach taken for Likert questions in the Australian Public Service Census.</p> <p><b>Additional rule if needed</b> Additionally apply the below '4' rule to individual responses or <a href="#">random rounding to base 3</a> if the suppressed values can be determined by looking across tables</p>
Sample surveys (e.g. Kiwis Count)	<p>Kiwis Count rules are about ensuring robustness of results, rather than confidentiality. The small size of the Kiwis Count survey means that it is very unlikely that an individual can be identified. Kiwis Count unit record data is published on the the Commission website after detailed regional council data is removed.</p> <ul style="list-style-type: none"> <li>• Results for questions with unweighted sample counts of less than 25 are suppressed with a 's' due to potentially high margins of error.</li> <li>• Results for questions with unweighted sample counts of less than 50 (i.e. those answered by less than 50 people) are released but are flagged that they may be subject to high margins of error due to small samples.</li> </ul>		

\* A likert question is one where someone is asked to rate their view on topic on scale (e.g. 5-point scale from strongly agree to strongly disagree). Typically, it is not possible to identify people from their response to a likert question (e.g. whether they were satisfied with their work-life balance). If it is possible (e.g. how much te reo they use at work, then suppress cells with counts below '4' as well.

**Table 2 – the Commission Data Release Rules for Magnitude Medians and Percentiles**

<b>Percentile</b>	<b>Minimum count needed overall</b>
1 <sup>st</sup>	500
5 <sup>th</sup>	100
10 <sup>th</sup>	50
25 <sup>th</sup>	20
50 <sup>th</sup> (median)	10
75 <sup>th</sup>	20
90 <sup>th</sup>	50
95 <sup>th</sup>	100
99 <sup>th</sup>	500

**Table 3 – Workforce Data person-level variables: Personal Data?**

<b>Variable</b>	<b>Personal data?</b>
Gender	No for Male or Female (personal, but not private) But, Yes for Another Gender
Ethnicity	Yes
Date of Birth / Age	Yes
Status – On Parental Leave / On other LWOP	Yes
Salary	Yes
Termination reason	Yes
Sick and domestic leave taken	Yes
Occupation / Job Title	No
Department / Business Unit	No
Status – Current employee / Seconded / Previous employee	No
Full-time Equivalent / Part-time	No
Contract Term (Permanent or fixed-term)	No
Tenure / Start or end date	No
Region	No
Management tier	No