





































- 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.
- Security – is data sensitive? If yes, more mature methods are desirable.
- Turnaround time – Is speed of collection and analysis important? If yes, more mature methods are desirable.
- Validating the data we collect:
  - When information is collected the data needs to be validated to ensure:
    - The data has been collected as intended and is consistent with other instances of the collection (if the information is collected regularly).
    - 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 information is analysed such as:
    - Changes over time – if there are large differences this may be due to an error in the data; otherwise determine the reasons for the change.
    - Large differences compared with other respondents.

It is also important to note if any information is going to be released to the public (regardless of where the collection is managed) needs to be checked and the [Information Release Guidelines](#) should be adhered to. – SIT complete checks to ensure the information we release is accurate and confidential.

## 8. Collections are reviewed regularly:

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

### Related Guidance:

[Information Strategy Diagram](#)

[Information Management & Release Policy](#)

[Information Management Protocols](#)

[Information Release Guidelines](#)

[Information Confidentiality Guidelines](#)

## Appendix 1. Collection Channels

### More mature channels:

- Secure portals that require data submission:
  - Survey Monkey - Surveys are created in Survey Monkey and the data captured from respondents is securely stored on the Survey Monkey server. The data is owned and administered by the survey creator, however Survey Monkey do have the right to access the survey data to provide the survey creator and respondents support. Survey Monkey also have the right to share the information you capture in some cases (such as if they obtain your consent or if they aggregate or de-identify the information).

Survey Monkey allows the survey creator to control who completes the survey. Survey Monkey is a good option for collecting information from a number of individuals based on specific questions identified to answer an information 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.
  - Next Cloud - Next cloud can be used for sharing information by loading the data into next cloud and setting up access for users you want to share the information 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 HRC 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 information.

Channels such as CFISnet and APIs would only be used for Strategic Information and managed by SIT.



## Information Release Confidentiality Guidelines

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

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

The guidelines are consistent with Privacy Act requirements and were developed in consultation with Stats NZ (see Stats NZ's Data [Confidentiality report](#) for more detail).

The guidelines apply to:

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

### 1. Roles and responsibilities are adhered to

- All staff have a duty to take reasonable care of sensitive information provided to them, this can be physically, electronically or even verbally.
- TKM staff should know their [responsibilities](#) with TKM information and their obligations to protect confidentiality and privacy.
- Prior to being granted access to sensitive strategic information, staff need to acknowledge the rules set by the data steward.

### 2. Use of confidential data within TKM

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

### 3. Information release rules

- All users of information should follow the release rules provided in Tables 1 and 2 below. The release rules provide statistical advice on aggregation, suppression and counts.
- Information release rules apply to all personal information, which are outlined in table 3.
- TKM can choose, in the following circumstances, not to apply the rules:
  - When supplying information back to the original supplier of the data. For example, agency workforce information back to the HR team of that agency.

- If information is already public or can be easily ascertained by a member of the public. For example, the gender of senior leadership team members.
- When reporting information 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 information.
- **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 information not being statistically robust according to Stats NZ.
- Note that Te Kawa Mataaho has an agreement with GCSB and NZSIS to not release any information collected in the Workforce Data, beyond that published in these [three tables](#).

**Table 1 – TKM Information Release Rules**

	<b>Counts (e.g. Headcount, FTEs)</b>	<b>Magnitudes (e.g. Salary, Sick Leave) Means &amp; totals</b>	<b>Magnitudes (e.g. Salary, Sick Leave) Medians &amp; percentiles</b>
Full coverage (e.g. Workforce Data collection)	<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><b>Or</b></p> <p>Apply random rounding to base 3 by using this <a href="#">workbook</a>. Note any use of random rounding. Totals and percentages should be calculated using rounded numbers.</p> <p><b>Additional rule if needed</b></p> <p>Apply both rules if many tables are being produced that may mean values can be determined by looking across tables or a highly sensitive variable is involved (e.g. income).</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> <p>Round magnitude values to an appropriate level. For example, round average salary to the nearest \$100, round average tenure, age and sick leave to one decimal place. Percentages should be calculated using rounded numbers.</p> <p><b>Additional rule if needed</b></p> <p>Suppress cells with counts below '10' in the following circumstances:</p> <p>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 information is high and the value is highly sensitive (e.g. income).</p>	Apply table 2.

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 TKM website after detailed regional council information is removed.</p> <p>Results for questions with unweighted sample counts of less than 25 are suppressed with a 's' due to potentially high margins of error.</p> <p>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.</p>
-----------------------------------	--

**Table 2 – TKM Information Release Rules for Magnitude Medians and Percentiles**

Percentile	Minimum count needed overall
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 Information?**

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