# PROCESS FOR PUBLISHING SENSITIVE UNIT RECORD LEVEL PUBLIC DATA AS OPEN DATA

**Objective:**
As set out in the Australian Government Public Data Policy Statement, data held by the Australian Government is an important strategic national resource that holds considerable value for growing the economy, improving service delivery and transforming public policy outcomes. The Australian Government also recognises that the privacy of individuals is of paramount importance. The purpose of this process is to set out a method to release ***sensitive unit record datasets*** as ***open data***—to ensure that these datasets have been protected to the highest standard.

**Scope:**
The process only applies to the release of sensitive unit record datasets as open data. The process is not intended to cover:

- unit record datasets that are not sensitive;
- datasets that are too sensitive to be made openly available, and cannot be adequately de-identified; or
- datasets that are only released by the data custodian to a limited number of researchers under contract.

Figure 1 sets out a process for the data custodian to determine whether the process applies to their dataset.

**The process:**
The following step-by-step process must be undertaken by the ***data custodian*** to release a sensitive unit record dataset as open data. Figure 2 illustrates the process.

1. The data custodian has determined that a sensitive unit record dataset is suitable to be published as open data (Figure 1 refers).

2. The data custodian informs the Office of the Information Commissioner (OAIC) (enquiries@oaic.gov.au) and the Department of the Prime Minister and Cabinet's (PM&C) Public Data Branch (data@pmc.gov.au) of its intention to release the sensitive unit record dataset and an estimated timeframe for publication.

3. The data custodian determines a methodology to confidentialise the dataset. Where a data custodian has internal ***data privacy experts***, the methodology may be developed internally. Data custodians that do not have sufficient internal expertise must engage an external data privacy expert. This could include entering into an arrangement with another Government agency that has data privacy experts. Alternatively, a data custodian could enter into an agreement with data privacy experts in academia or the private sector.

4. The data custodian must engage a different data privacy expert to review the methodology developed in item 3. The second data privacy expert's review needs to take into account:
   a. the primary audience for the dataset, and its uses;
   b. whether the methodology to confidentialise the sensitive unit record dataset is sufficient to meet the required levels of de-identification in the dataset. If the second data privacy expert considers that the methodology to confidentialise the sensitive unit record dataset will never be sufficient to meet the required levels of de-identification in the dataset, the sensitive unit record dataset will not be released as open data; and

c. advances in technology. Methods that were sufficient to de-identify data in the past may become susceptible to re-identification in the future. The second data privacy expert needs to ensure the methodology takes into account advances in technology, as well as establishing a process to periodically review the confidentialisation methodologies.

The same data privacy expert cannot be engaged for the review that developed the methodology in item 3. For example, if a data custodian engages an external data privacy expert to determine the methodology to confidentialise the dataset in item 3, a different external data privacy expert must be engaged for this item. If the data custodian engaged internal data privacy experts in item 3, it must engage an external data privacy expert for this item.

4.1 An optional augmentation to item 4 is to publish a high-level methodology to confidentialise the sensitive unit record dataset publicly prior to the release of the dataset. This additional step allows a high level confidentialisation methodology to be publicly tested, and any identified weaknesses corrected.

5. A public report will be generated by the data custodian, setting out:
   a. the confidentiality measures used to de-identify the dataset;
   b. the outcomes of the consultations with the data privacy experts, and if applicable, the public consultation on the methodology;
   c. how the data custodian has taken account of the Australian Privacy Principles, and the Office of the Australian Information Commissioner's (OAIC) Guide to big data and the Australian Privacy Principles (currently in draft); and
   d. a data dictionary, which explains the fields in the dataset.

6. The ***data custodian Agency-Head***:
   a. is provided a copy of the public report; and
   b. considers approving the release of the sensitive unit record dataset as open data.

By approving the release of the dataset, the data custodian Agency-Head is taking responsibility for the confidentiality measures used to de-identify or confidentialise the sensitive unit record dataset. In taking this responsibility, the data custodian Agency-Head needs to ensure that appropriate data privacy experts, with relevant skills, have been engaged to advise on the methodology to confidentialise the sensitive unit record dataset, and that the advice received as a result of items 3 and 4 have been adequately addressed.

7. For their information, the public report will be provided by the data custodian to:
   a. the Minister(s) with responsibility for the publishing agency; and
   b. PM&C's Public Data Branch (data@pmc.gov.au), who will also provide the public report to the Office of the Special Adviser to the Prime Minister on Cyber Security and the Assistant Minister for Cities and Digital Transformation.

At their next respective meetings, the Secretaries and the Deputy Secretaries Data Groups will also be informed about the release of the dataset.

8. Once the data custodian Agency-Head has approved the release of the sensitive unit record dataset as open data, the data custodian can publish the dataset on data.gov.au, with a copy of the public report, and the data custodian's contact details.

9. In the event of a potential vulnerability being reported to the data custodian, PM&C, or another Government entity, the PM&C Public Data Branch will immediately mark the sensitive unit record dataset as 'private', removing it from public view. The data custodian will then be informed of the

potential vulnerability, who will then subsequently inform the OAIC. The data custodian will then undertake an investigation into the potential vulnerability. The objective of the investigation will be to consider whether items 3–8 were adequately implemented, and whether the dataset can be re-published once the potential vulnerability has been addressed.

**Definitions:**
***Sensitive unit record dataset*** is a dataset that includes unit record level data, such as personal or commercial-in-confidence information.

***Sensitive*** means data that can be used to identify an individual, species, object, or location that introduces a risk of discrimination, harm, or unwanted attention. Common categories of sensitive data include human medical, health, and personal data.

***Open data*** is data that is accessible to everyone to download and utilise. This includes datasets that are available on data.gov.au and/or a data custodian's website without secure access or a deed.

***Data custodian*** is the Australian Government entity, or entities, responsible for the dataset.

***Data privacy experts*** are personnel with particular skills and expertise in data de-identification approaches and re-identification attack risks.

***Data custodian Agency-Head*** is the person responsible for releasing the sensitive unit record dataset as open data. The data custodian Agency-Head may delegate this responsibility to another person, or to a committee where appropriate.

**Figure 1** sets out a process for the data custodian to determine whether the process applies to their dataset.

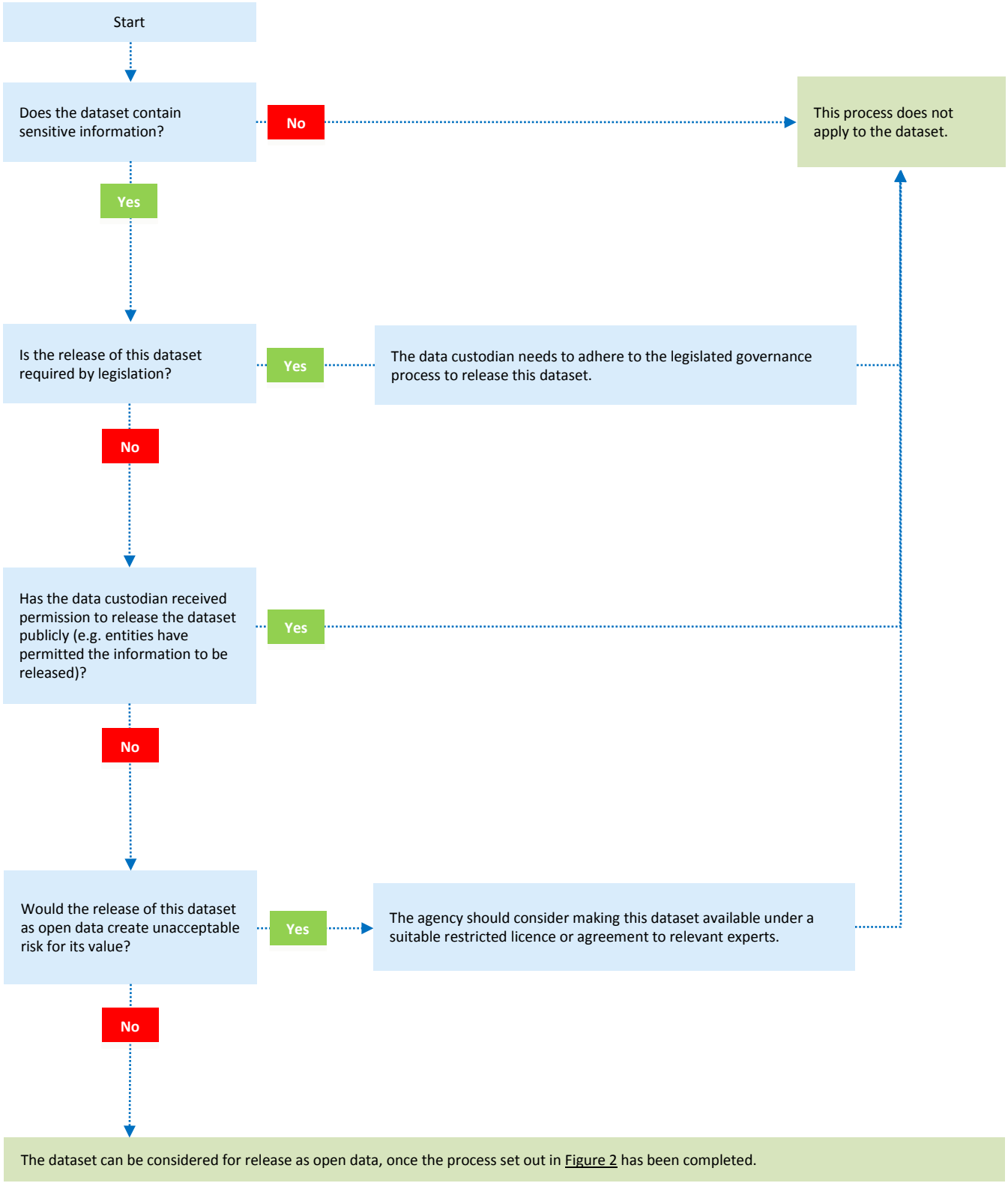

**Figure 2** is a step-by-step process which must be undertaken by the data custodian to release a sensitive unit record dataset as open data.

Step 1: The data custodian has determined that a sensitive unit record dataset is suitable to be published as open data (Figure 1 refers).

Step 2: The data custodian informs the OAIC and PM&C's Public Data Branch with an estimated timeframe for the release of the data.

Step 3: The data custodian determines a methodology to confidentialise the dataset. Depending on the data custodian's capability, the methodology may be developed internally, or externally.

Step 4: The data custodian will engage a second data privacy expert to review the methodology. The review needs to take into account:

a. the primary audience for the dataset, and its uses;
b. whether the methodology to confidentialise the sensitive unit record dataset is sufficient to meet the required levels of de-identification in the dataset; and
c. whether the methodology takes into account advances in technology, as well as establishing a process to periodically review the confidentialisation methodologies.

Step 4.1: An optional augmentation to item 4 is to publish a high level methodology to confidentialise the sensitive unit record dataset publicly prior to the release of the dataset. This additional step allows the high level confidentialisation methodology to build public confidence in data release and make transparent any shortfalls that can be addressed.

Step 5: A public report will be generated by the data custodian. The draft public report will include the confidentiality measures used to de-identify the dataset, and the outcomes of the data privacy expert review.

Step 6: The data custodian Agency-Head:

a. is provided a copy of the public report; and
b. considers approving the release of the sensitive unit record dataset as open data.

By approving the release of the dataset, the data custodian Agency-Head is taking responsibility for the confidentiality measures used to de-identify the sensitive unit record dataset.

Step 7: For their information, the public report will be provided by the data custodian to:

a. the Minister(s) with responsibility for the publishing agency; and
b. PM&C's Public Data Branch, who will also provide the draft public report to the Office of the Special Adviser to the Prime Minister on Cyber Security and the Assistant Minister for Cities and Digital Transformation.

At their next respective meetings, the Secretaries and the Deputy Secretaries Data Groups will also be informed about the release of the dataset.

Step 8: The data custodian can publish the dataset on data.gov.au if the Agency-Head has approved the release of the dataset as open data.

Step 9: In the event of a potential vulnerability being reported to the dataset custodian, PM&C, or another Government entity, the PM&C Public Data Branch will immediately mark the dataset 'private', removing it from public view. The data custodian will then be informed of the potential vulnerability, who will then subsequently inform the OAIC. The data custodian will then undertake an investigation into the potential vulnerability.