NZTA’s open data - framework and processes 2019

Defining how we think about, and treat, our open data

aimee whitcroft, Business Intelligence SERVICES

29 October 2019

V1.3

**This document is licensed by NZTA under a CC-BY 4.0 International licence.**

<https://creativecommons.org/licenses/by/4.0/>

What is open data? 2

Why is it important? 2

For NZTA 2

For our customers, stakeholders and all New Zealanders 2

What are our obligations? 3

Why do we need an open data framework? 3

Governance is key 3

Our open data framework 4

Overview 4

When you update a dataset 5

Identify 5

Assess 6

Prioritise 6

Prepare 7

Approve 7

Release and maintain 8

Measure and report 8

Important considerations 9

Open data standards and formats 9

Metadata 9

Data quality 10

Procurement 10

# What is open data?

Open data is defined as “data anyone can use and share. It has an [open licence](https://www.data.govt.nz/manage-data/policies/open-licensing/), is openly accessible and is both [human-readable](http://opendatahandbook.org/glossary/en/terms/human-readable/) and [machine-readable](http://opendatahandbook.org/glossary/en/terms/machine-readable/).”[[1]](#footnote-1)

Open *government* data is non-personal, unclassified and non-confidential.[[2]](#footnote-2)

## Why is it important?

Open data isn’t an end in itself, and doesn’t (shouldn’t) exist in isolation in an organisation.

Instead, it should be wound throughout the processes and systems of the organisation, meeting the needs of the people both within and outside her.

It’s both a lever for and an output of data management and governance processes, and a strong indicator of an organisation’s data maturity.

### For NZTA

* To meet our statutory obligations.
* To help us meet our strategic objectives[[3]](#footnote-3).
* To help build more innovative products and services.
* To help enable better data sharing and use within the organisation – opening data helps organisations use their own data by breaking down silos.
* To help us interact better with our customers and stakeholders, improving our user focus, collaboration efforts, reputation and social licence.

### For our customers, stakeholders and all New Zealanders

* To help build a safer, more inclusive, more efficient land transport sector in NZ.
* To help build data capability and capacity throughout the land transport sector and related stakeholders.
* To help empower New Zealanders in their land transport uses, both as users and as organisations which provide products and services.
* To help improve land transport-related decision-making, in turn leading to improved funding – and more effective use of funding – and thus a wide range of fundable activities and initiatives.

The lists above aren’t exhaustive, but give some idea of the tremendous power of better sharing and opening data.

If you’re interested in getting involved with public-facing open data communities around New Zealand, there are a number of open data meetups as well as the Open Government Ninjas mailing list and forum[[4]](#footnote-4).

There’s also the ‘open data practitioners’ monthly lunchtime session open to people actively working on open data inside and outside government[[5]](#footnote-5).

## What are our obligations?

NZTA is a signatory to the New Zealand Data and Information Management principles[[6]](#footnote-6). These state that data should be:

* open
* protected
* readily available
* trusted and authoritative
* well-managed
* reasonably priced (preferably free)
* reusable.

Further, the principles state that “unless there are identifiable reasons for its release, personal and classified data remains protected. Government data and information should also be trusted and authoritative.”[[7]](#footnote-7)

New Zealand is also signed up to the Open Data Charter, a set of international principles. Of these, “publish with purpose” helps to balance the “open by default” approach laid out above, as it helps agencies identify which datasets to open first / focus most effort on.

Finally, The Ministry of Transport’s Domain Plan – currently being updated – also identifies open data as a priority for the sector[[8]](#footnote-8).

# Why do we need an open data framework?

NZTA has developed this open data framework and supporting materials to bring together our open data work in a cohesive fashion.

# Governance is key

In any system, governance is key – doubly so when it comes to data, which carries huge potential but also, especially when involving personal information, risk.

In this document, we define governance as the entire set of capabilities and roles which ensure that data is properly managed and maintained. We don’t simply mean the Board.

We’re in the process of defining how these various roles work, and will link to them when they’re ready. For the purposes of this document, the primary roles are:

* data custodian(s) – people at the top of the organisation responsible for setting enterpri-se policities, standards and rules, as acting as an executive sponsor for data-related work and initiatives
* data stewards – people responsible for the day-to-day management of datasets
* business owners – people responsible for the content of the datasets, and understanding their proper use
* subject matter experts – these could be data stewards, business owners or other people with a detailed understanding of various aspects of a dataset
* data coordinator – people responsible for helping to triage, route and coordinate data activities – especially related to shared and open data – through the organisation.

# Our open data framework

## Overview

*Image: diagram of the open data framework and related data governance layers. For text version, see sections below. Credit: NZTA, CC-BY 4.0 International.*

The steps below explain what to do at each point in the 7 steps of the framework.

Primary documents:

* *open data toolkit 2019* – this contains most of the documentation you’ll need for preparation and approval.
* *master open data directory* – this is where you’ll store the details of open datasets, requests and feedback.

You can also refer to the ‘open data process – intro’ tab in the *open data toolkit, 2019*, for a summary reminder of the guidance below.

We also recommend you read the preceding tab in the same workbook, called ‘data considerations – overview’. These are useful points to bear in mind as you work with data.

Every time you begin the process of opening a new dataset, you must:

* make a copy of the *open data toolkit, 2019*
* place it in the *completed open dataset documentation* folder in InfoHub

We expect that either the business owner or data steward will be the people filling in the toolkit – except for approvals – as it goes through the process.

Should NZTA define a “data coordinator” role, that person / those people will help fill out the appropriate documentation and guide people through the preparation, publishing and maintenance processes.

### When you update a dataset

Note: not all datasets will require each part of the framework, especially with datasets which are already published. In those cases, choose the appropriate phases to ensure proper governance, maintenance and reporting.

Ensure you update the *master data directory* whenever you update a dataset, so it reflects the most up to date information.

For a published dataset, you only need to complete an open data toolkit 2019 if:

* it’s a major re-release
* it’s the product of a request, and needs to be prioritised.

## Identify

This first step of the framework is all about identifying datasets which are (possible) candidates for opening - these could be identified internally, or as a result of external requests.

Possible sources of candidate dataset suggestions:

* user groups
* external requests (eg via data.govt.nz, Official Information Act requests, etc)
* internal requests.

**Action**

* **record all dataset requests in the requests log in the *master open data directory*.**

For the rest of this document, we’ll take about a single example dataset which members of the public have requested be opened.

People involved:

* data steward
* data coordinator (?).

## Assess

Now that you’ve recorded the request in the requests log of the *master open data directory*, it’s time to begin the assessment process.

This toolkit does not identify whether you need to conduct further assessments, such as privacy, risk, etc. These should be part of your team’s normal business processes.

To get initial sign-off to continue (see end of sheet), you must have:

* completed the sheet
* completed all necessary assessments.

The person signing off should be at an appropriate level of knowledge and accountability to do so.

We recommend your team manager as a good person for this.

If you are unable to get sign-off to continue the process, make a note in the worksheet in column D explaining why, and stop the opening process.

**Actions**

* **Complete any necessary assessments (privacy, etc).**
* **Make a copy of the *open data toolkit 2019*.**
* **Name the copy with the name in the request log, and place the copy in the *completed open dataset documentation* folder.**
* **Complete the ‘data discovery’ tab in your copy of the toolkit.**

People involved:

* data steward
* business owner’s or data steward’s manager
* data coordinator (?).

## Prioritise

Once you have sign-off to continue, the next step is to prioritise the dataset for opening. Over time, as we look at opening more data, we will have a growing backlog of potential datasets.

This simple process will assign a priority to the dataset. You can then communicate this priority:

* to the initial requesters
* to your users groups and the wider community (if appropriate)
* to NZTA, as needed.

The prioritisation process takes into account all-of-government, Ministry of Transport and NZTA objectives. It also takes into account the amount of effort opening the dataset will involve.

**Actions**

* **Complete the ‘release prioritisation’ tab in your copy of the *open data toolkit 2019*.**
* **Update the requests log in the *master open data directory* with the ‘final prioritisation ranking (board 1)’ score at the bottom of the prioritisation tab.**
  + **This will be a single number between 0 and 16.**
* **If there are any other considerations:**
  + **record them as an email / emails in the *completed open dataset documentation folder*, and**
  + **link to the email(s) in Column D of the prioritisation tab in your copy of the open data toolkit.**

People involved:

* data steward
* subject matter expert (if necessary)
* data coordinator (?).

## Prepare

Working from the most to the least important datasets in terms of prioritisation is vital. Once a dataset’s turn has come up for release, it’s time to prepare the dataset and related metadata for final sign-off to release.

If there are any reasons to accelerate a dataset’s release:

* record them as an email in the completed open datasets folder
* link to that email in Column D of the ‘final release signoff’ tab in your copy of the open data toolkit.

**Actions**

* **Complete the ‘data directory output’ tab in your copy of the *open data toolkit 2019*.**
  + **In some cases, this can only be fully completed once the dataset has been published.**
* **Prepare the dataset itself for publishing, including any anonymisation required (as identified in the risk / privacy assessments linked to – if they exist – in the ‘data discovery’ tab).**
* **Capture any other metadata and documentation (eg methodology, caveats, licence detail) required, in 'completed open dataset documentation' folder in InfoHub.**

People involved:

* data steward
* business owner
* other subject matter experts as required to prepare the dataset
* data coordinator (?).

## Approve

Once everything’s ready, it’s time to get final sign-off.

You may wish to include a memo or further information to the those signing-off, for example briefly capturing what the user demand is for opening the data, or what issues opening it will solve.

Be aware that sign-off can take a while, so don’t leave this to the last minute if there’s a hard deadline.

**Actions**

* **Complete the ‘final release signoff’ tab in your copy of the *open data toolkit 2019*.**
  + **If you can't get sign-off - see end of sheet - STOP HERE.**

People involved:

* data steward and business owners’ managers
* data coordinator (?)
* data custodian (?).

## Release and maintain

If you’ve been successful in getting sign-off, you can publish the dataset(s) on our open data portal.

This means you’ll need to coordinate with the team managing the portal. As with sign-off, be aware this may not be possible to do instantly, as the team will have other priorities, too, into which you’ll need to fit this work.

Having everything fully prepared and signed off will help everything publish / update the datasets quicker and more easily.

You’ll also need to update our internal register(s) with the relevant details about the release, to help ensure we’re tracking and updating everything efficiently.

Finally, you’ll need to coordinate with a number of teams in the organisation to ensure that the relevant people in and outside the organisation know about the dataset’s release.

**Actions**

* **Update the 'open data directory' in the *master open data directory*.**
* **Publish / update the dataset on the open data portal.**
* **If necessary, publish / update the dataset's listing on data.govt.nz (note: this may be done automatically via the ODP)**
  + **https://www.data.govt.nz/**
* **Alert comms, contact centre and data services with new data details, and whom to contact for questions**
* **Alert user groups with new data details.**

People involved:

* data steward
* data coordinator (?)
* content uploaders for open data portal.

## Measure and report

While this is the final step in the overall open data framework, this should be an ongoing process woven throughout all our other open data activities.

It’s vital to measure and report against our open data to ensure that it’s reaching the right people and having the desired effect.

Good measurement and reporting will also help ensure our open data release and maintenance efforts are properly prioritised and supported throughout the organisation, and carried out with our users – internal and external – front and centre in our minds. It’s an important part of the “publish with purpose”[[9]](#footnote-9) principle to which NZ’s government agencies have signed up.

**Actions**

* **Measure and report against open data use and feedback, capturing reports in the 'reports log' of the *master open data directory*.**
* **Capture feedback about existing open data in the 'requests log' of the *master open data directory*.**
* **Ensure feedback and reports are sent to appropriate people in organisation, for use in improving datasets and priorities.**

People involved:

* data steward
* business owner / SME
* data coordinator (?).

This last step – measure and report – will feed information back into earlier steps, including, but not limited to:

* identify (people asking for new datasets, or new data in existing open datasets)
* prioritise.

# Important considerations

## Open data standards and formats

It’s not enough simply to release data.

To help ensure people people can find and use it, you should release it:

* in an open data format, eg CSV, JSON, GeoJSON, KML, XML and RDF
  + PDF is *not* considered an open data format
  + XLS isn’t an open data format either, but if necessary is acceptable
* using open data standards
  + these are data standards which are available to anyone
  + there’s no definitive list of open data standards, but when releasing a dataset, check to see whether there are open data standards for the data – for example, the traffic asset management data standard[[10]](#footnote-10) is a set of open standards.

## Metadata

Data without good quality metadata is either invisible, or unuseable, or both.

Metadata tells people about a dataset – what it’s about, how to use (or not use!) the data, and its licence.

Ideally, metadata should also tell people how the data was collected, for what purpose, and what its quality is, as well as including a data dictionary.

## Data quality

If your data quality is poor internally, it will be poor externally. This can potentially pose a risk to the agency or our customers and stakeholders. It can also make it far more difficult to use the data, as people have to spend increased time cleaning and preparing the data.

However, ‘dirty’ (or lower-quality) data can still be publishable externally, if you make sure:

* its metadata is good
* it’s properly caveated
* quality dimensions are part of the metadata.

It’s important to release data as close to raw as you’re able to – the more cleaning and abstractions you add to, the more chance you’re hindering people’s ability to use it. However, you must balance this with privacy and risk considerations.

## Procurement

As you renew existing data agreements, or sign new ones with data providers, ensure that you bear in mind data ownership, control, and the ability to open data down the line.

This means that you should, wherever possible, only sign agreements with vendors where you become the owner or controller of:

* data resulting from the work you’re procuring
* any data itself that you’re procuring.

If these provisions aren’t specifically written into any contracts, vendors own the data, hamstringing your ability to put it to its most effective use down the line.

*End*

1. <https://www.data.govt.nz/open-data/what-is-open-data/> [↑](#footnote-ref-1)
2. Ibid. [↑](#footnote-ref-2)
3. “Position statements”, pgg 18-34, <https://www.nzta.govt.nz/assets/resources/statement-of-intent/2018-2022-amended/soi-2018-2022-amended.pdf> [PDF 4.5 MB] [↑](#footnote-ref-3)
4. <http://groups.open.org.nz/groups/ninja-talk> [↑](#footnote-ref-4)
5. For more details, contact aimee whitcroft@nzta.govt.nz. [↑](#footnote-ref-5)
6. <https://www.data.govt.nz/manage-data/policies/new-zealand-data-and-information-management-principles/> [↑](#footnote-ref-6)
7. <https://www.data.govt.nz/manage-data/policies/new-zealand-data-and-information-management-principles/> [↑](#footnote-ref-7)
8. Transport Domain Plan Full List of Recommendations, 2016 [↑](#footnote-ref-8)
9. [https://opendatacharter.net/principles](https://opendatacharter.net/principles/) [↑](#footnote-ref-9)
10. https://www.nzta.govt.nz/roads-and-rail/asset-management-data-standard/ [↑](#footnote-ref-10)