New dataset content checklist

The metadata describes the data and purpose of each dataset in the catalogue. It is collected through description forms and is used to populate the dataset webpages on Ontario.ca and form the government-wide data inventory dataset.

To help ensure the language is consistent and accessible to a wide audience of users here are some things to consider the next time you are tasked with creating or updating metadata for a dataset.

## Mandatory metadata requirements

*Ensure fields are filled out as required.*

* All mandatory fields are completed
* Optional fields completed where possible, left blank otherwise
* Descriptions are within the character limits
* All fields using a controlled vocabulary use one of the specified options
* Dates are full YYYY-MM-DD format, blank otherwise
* Topic is selected from the [list of official Ontario.ca topics](http://intra.cabinetoffice.gov.on.ca/communications/index.php/ontario-caonesite/ontario-ca-topics-and-tags-inventory/)
* Links to technical resources are included where relevant (e.g. Land Information Ontario metadata link), blank otherwise

## Plain language

*Ensure all fields are written using plain language. Evaluate against the criteria below.*

* Sentence case the titles) – e.g. “Government-wide data inventory”
* Capitalize only proper nouns, lower case ministry and minister when not used beside a ministry’s name or a minister’s name – e.g. The minister of the ministry approved this program. Yesterday, Minister Matthews approved a program for the Ministry of Advanced Education and Skills Development.
* Include what the user should expect to get out of this data set in the short description (lead)
* Long descriptions include more detail/background on the dataset and program
* Long descriptions don’t repeat the information in the short descriptions
* Consider structuring the long description:
  + To begin with “This data set contains…”
  + To include why / under what authority it was collected – e.g. “This data is collected as part of <your program>”
  + To describe what is in the dataset – e.g. “Fields include ministry name, …”
  + To include links to the program website
* List examples of what the data set contains in the long description
* Include periods after short and long descriptions

## Status

*The Open Data Directive requires ministry and agency data to be “Open by Default” unless one of the exemptions applies.*

* Datasets are marked “open” or “To be opened”, unless an exemption applies (“restricted”) or further review is required (“under review”)
* If dataset is restricted, ensure correct exemption is indicated
* If dataset is restricted, ensure rationale is included, otherwise leave blank
* Rationale should be plain language and indicate why the exemption applies to the dataset. E.g. “Dataset contains personal information about program recipients”
* Even if a dataset is restricted, consider if there are versions or views of the data that could be released. E.g. OPS HR data is restricted, but Workforce Demographics aggregate dataset is open data

## Due diligence

*Ensure the following due diligence steps have been taken.*

* Remove specific names of third party products or vendors – to avoid appearance of endorsement or taking credit for products. Use third party language as needed, file formats are fine. E.g. “third-party database” instead of “Oracle Database”
* Check if datasets are rolled up under any other ministries as corporate datasets, if yes they should be removed to avoid redundancy
* Ensure sets concerning indigenous population data were collected with proper consent