Building Datasets

Creating Metadata and Data Dictionary

Contents

1. Glossary

Serial	Term	Description/Explanation
1	**Metadata**	Metadata provides important structural and contextual information about the data; it describes characteristics and attributes of the data (e.g., who, what, where, why, how) . Metadata makes finding content and data faster and easier. Metadata facilitates data discovery and linkage across relevant and different data sources.
2	**Data Dictionary**	A Data Dictionary is the information you provide that tells a user exactly what is in your data resource and how it can be used

2. Creating Metadata and a Data Dictionary

2.1 What is Metadata and why is it useful and needed?

Metadata is information that you provide to describe your data resources. Search engines use metadata to help users find the data you put online. For the portal, data.govmu.org, the type of metadata is descriptive metadata to tell users what information the file contains (title, description, publisher, source file, URL concerned). The more information the search engine has about your data resources, the easier it will be for it to point to your resource when users search.

Metadata is what helps a user narrow down, for example, 400 data resources to just a few that will likely provide the data they need to answer their question.

To help in creating the metadata you will need for your data resource, a description of each Metadata field has been provided below:

Metadata Field	Required before publishing on data.govmu.org	Field Description	Metadata for Dataset
1. Title	Required	● Choose a name for your dataset that tells users what the topic of the dataset is.	
2. Description	Required	● This is a plain English description that will display below the name of the data table.	

		● Write a summary paragraph telling us what the data table contains.	
3. Tags		● Descriptive keywords or phrases that users will search for to find your data resources. ● Press enter after one tag and type another ● Try to include at least three descriptive tags. ● Both general and specific terms are useful.	
4. Groups	Required	● Also known as the publisher ● The Ministry, department, organisation, group, or parastatals that publishes the data resource. ● Choose from the drop down list.	
5. Topics	Required	● Also known as Category. ● Choose best one from the current list at data.govmu.org	
6. License	Required	● Use Creative Commons Attribution	
7. Author	Required	● The Ministry, department, organisation, group, or parastatals that authors the data resource and has ultimate responsibility for the creation of the data. ● Will often but not always be the publisher of the data.	
8. Spatial / Geographic Coverage	Optional	● The geographical area the data table covers. ● Specification should include a named area and may include geographic coordinates.	
9. Frequency	Required	● How often do you intend to publish or update the data resource on data.govmu.org? ● E.g. Annually, quarterly, monthly	

10. Temporal Coverage	Required	● Start date and End date for the data in your data resource.	
11. Granularity	Required	● Tell us the most specific data that your resource gets. ● Often measured in geography (census track) or time (monthly or daily data).	
12. Data Dictionary Type	Required	● If the link below is not an HTML file, the file type for the data dictionary.	
13. Data Dictionary	Required	● HTML link to the data dictionary itself.	
14. Contact Name	Required	● Enter Open Data Team contact	
15. Contact Email	Required	● Enter a generic e-mail address for your Open Data Team or group	
16. Public Access Level	Required	● Whether this info could ever be made public. (Public, Restricted, Non-Public)	Public
17. Rights	Required	● If you entered anything but Public above, you must explain any use restriction on the data. ● Additionally, you can also include usage/research/collaboration instructions: Short text or link to a document that describes how the data can be used, research Ideas and/or possible collaborations based on this information that may interest external researchers. ● If not applicable, please enter "No restrictions on public use".	
18. Homepage URL	Optional	● URL for the page on your website that has useful information about the data resource or the group that updates it.	
19. Data Standard	Optional	● A technical description of the data	

		● E.g. CSV, XML standards, or JSON.	
20. Language	Optional	● Most often English.	
21. Additional Information	Required	● Additional information is a field that allows you to enter free form metadata in key-value pairs: a key, which is a unique identifier for some item of data, and the value, which is the data that is identified. ● You must include at least two keys-value pairs: one with the "Limitations" key and one with the "Data_Methodology" key. : ● Data Methodology: Short text or link to a document with explanation of the data collection methodology, which may include survey tools, post-collection methods for control and cleaning and notes on sampling response and errors. This is where you should document any known issues in using the data. May also be uploaded as an additional PDF.	
22. Resources	Required	● URL for the data resource web page for further information on the dataset, its distributions and any additional information.	
23. Related Content	Required	● Enter secondary source(s) info: If your data resource is partially made from other data sources, please give descriptive name(s), and/or URLs, of resource(s) from which the data table is derived.	

2.2 What is a Data Dictionary and why is it useful and needed?

A Data Dictionary is the information you provide that tells a user exactly what is in your data resource and how it can be used. It lists all the variables, both by the name they go by in the resource as well as a plain English name, the format (text, numeric, date) for each, and a description of what information the user can find in that variable.

A Data Dictionary is what helps the user decide if your data resource will provide the data that is needed to answer their question, and also if it is in a form they can use.

To create the data dictionary for your resource, the table below may be used; save the file as a PDF and upload it as an associated resource with your dataset.

Dataset Name:

Field Title	Field Name	Data type	Description

Field Title: Common English title for the data contained in this column. Please avoid abbreviations if possible.

Field name: The name of the field from the data table

Data Type: Choose one for each named column:

- Plain Text
- Formatted Text (allows bold, italics, etc.)
- Number (only use "number" if your data is meant to be added, otherwise use "plain text")
- Money
- Percent
- Dates/Time
- Location

Description: Full description of the values included in the column. If the column is a category, such as age group, then all categories or levels should be listed.

3. Next

The next step is getting approval for publishing from your Designated Officer and upon approval, upload your dataset.