

DKAN Quick Start Guide

Adding a Dataset

Information in this document contains excerpts from the comprehensive DKAN manual available at <https://docs.getdkan.com/en/latest>

Introduction	1
Step 1: Create Dataset	3
Access Control and Licensing	4
Step 2: Adding Resources	4
CSV Format - Comma Separated Values	6
GeoJSON Format	6
Step 3: Additional metadata	8
Advice for Working With CSV Files	9
Formatting data as CSV	9
Handy checklist	9

Introduction

Datasets are “Containers” that hold resources. Datasets are given metadata such as author, license type etc, and they can be assigned to Groups. One dataset record may contain multiple resource records, which are used to provide multiple file formats, data variations or related information such as reports.

Nauru Sea Levels ← Dataset title

Coastal and Marine
Sea level data by month and year 1993 to 2017

Group (dataset provider) → Department of Commerce, Industry and Environment
The Nauru Department of Commerce, Industry and Environment has oversight and responsibility over Issues concerning the environment, climate change, and commerce.

Multiple resources (downloads)

- Sea levels by month [Preview] [Download]
- Sea levels by year [Preview] [Download]
- [Download All]

Metadata (data about the dataset)

Field	Value
Publisher	Department of Commerce, Industry and Environment
Modified Date	2018-02-02
Release Date	2018-02-02
Frequency	Monthly
Homepage URL	http://www.bom.gov.au/ntc/IDO70055/IDO70055SLD.shtml
Identifier	a5a1f3f1-4f0d-47ef-a358-a4a671f18a34
Spatial / Geographical Coverage Area	POLYGON ((166.7374420166 -0.65285700418652, 166.7374420166 -0.44275319999295, 167.0890045166 -0.44275319999295, 167.0890045166 -0.65285700418652))
Spatial / Geographical Coverage Location	Nauru
Temporal Coverage	Friday, January 1, 1993 - 00:00 to Monday, December 31, 2018 - 00:00

Data Extent

License (usage restrictions) → Other (Public Domain)

Step 1: Create Dataset

From the Admin Menu, hover over the Add Content menu link. On the drop-down menu, select the Dataset option.



Add content -> dataset

Add values for the basic metadata fields such as title, description, tags, groups and topics. Specify a license to clarify reuse limitations. Click 'Next: add data' to continue

A screenshot of a web form titled 'Add a Dataset'. The form is divided into three steps: '1 Create dataset' (highlighted in green), '2 Add data', and '3 Additional data'. The 'Create dataset' step contains several fields: 'Title *' with a placeholder 'eg. A descriptive title'; 'URL' with a placeholder 'dkan.nauru.eightyoptions.com.au/dataset/[dataset]' and an 'Edit' link; 'Description' with a rich text editor toolbar and a placeholder 'eg. Some useful notes about the data'; 'Text format' with a dropdown menu set to 'Markdown HTML' and a link 'More information about text formats ?'; 'Tags' with a text input field and a placeholder 'eg. economy, mental health, government'; and 'Topics' which is partially visible at the bottom.

Add Title description and tags

Access Control and Licensing

All metadata describing a dataset and publishing the existence of a dataset will be publicly available. However, access to the actual dataset resource download files can be restricted using the *License Options and Access Control* field.

License Options and Access Control *

- Select a value -

- **Public domain** - Dataset resources will be freely available for download and use without restriction.
- **Creative Commons Attribution Share-alike** - Dataset resources will be freely available. Usage must give attribution and share the dataset under the same license.
- **Inter-agency** - Metadata will be publicly available but actual dataset resources will only be available for download to authenticated users.
- **Private** - Metadata will be publicly available but actual dataset resources will only be available for download to users within the same group.

Step 2: Adding Resources

Resources are the actual files or references for the dataset being published. Multiple resources can be added to one dataset, which can be useful for publishing multiple formats or variations of the dataset. Resources can also be used to publish other related references such as a copy of a report that a dataset was used for.

Three methods are provided for adding a reference file:

Upload

Drag and drop a file into the upload area or select *Browse* to choose the file to upload.

Embed Resource

Provide a URL to a file provided on an external website and the data portal will attempt to embed and display that resource directly on the page. This is useful when referencing external images so that they are visible on the resource page directly in the data portal. Note however that some website block their resources from being embedded..

Link to Remote File

Provide URL link to resource provided by an external website. The link will be published as a standard hyperlink.

Add a Resource

1 Edit dataset 2 Add data 3 Additional data

Upload Embed Resource Link to Remote File

Drop a file here or click *Browse* below.

[Browse](#) Upload

Files must be less than **10 GB**.
Allowed file types: **csv html xls json xlsx doc docx rdf txt jpg png gif tiff pdf odf ods odt tsv tab geojson xml zip.**

▾ Data Previews

Select desired previews for this file.

Grid

Graph

Map

Delimiter

▾

Select delimiter for file if applicable.

Embed

Provide an embed link?

Title *

Description

Add a Resource

If providing a data format that DKAN can parse (CSV, GeoJSON), DKAN can generate grid, graph or map visualisations of the dataset. Enable the desired data preview options where appropriate for your dataset.

Provide a title and description for the dataset. If providing multiple resources for one dataset, it is a good idea to include in the title how the resources are different. E.g.

- 'Rainfall data - Monthly'

- 'Rainfall data - Yearly'

After completing the fields on a resource, select *Save* to save the view the resource page, *Save and add another* to add another resource record to the same dataset or *Next: additional* information to continue to step 3 where more metadata can be provided.



Additional resources can be added to a dataset at a later point in time by navigating to the dataset and selecting *Add Resource*.



Nauru Sea Levels

Coastal and Marine

CSV Format - Comma Separated Values

CSV format is a common, simple format for producing a machine-readable spreadsheet of data. Wherever possible, it is recommended that published datasets are made available in CSV format. DKAN is capable of reading CSV formatted data and provides visualisation options for generating tables, graphs and maps, making the dataset easily accessible and browsable to users. See ***Advice for Working with CSV Files*** at the bottom of this document for further details on working with CSV files.

Data from CSV files can also be imported into the DKAN Datastore. Doing this allows the data to be queried via the DKAN APIs, allowing the data to be integrated directly into other applications such as for generating charts.

Import all content from source?

[View](#) [Edit](#) [Revisions](#) [Back to dataset](#) [Manage Datastore](#) [Download](#) [Data API](#)

[Import](#) [Delete items](#) [Unlock](#) [Drop Datastore](#)

DKAN Datastore File: Status

- No imported items.
- Next import: not scheduled.

Delimiter

The character that delimits fields in the CSV file.

No Headers

Check if the imported CSV file does not start with a header row. If checked, mapping sources must be named '0', '1', '2' etc.

File encoding

Performs character encoding conversion from selected option to UTF-8.

Geolocate

Geocoding requires enabling the Geocoder module

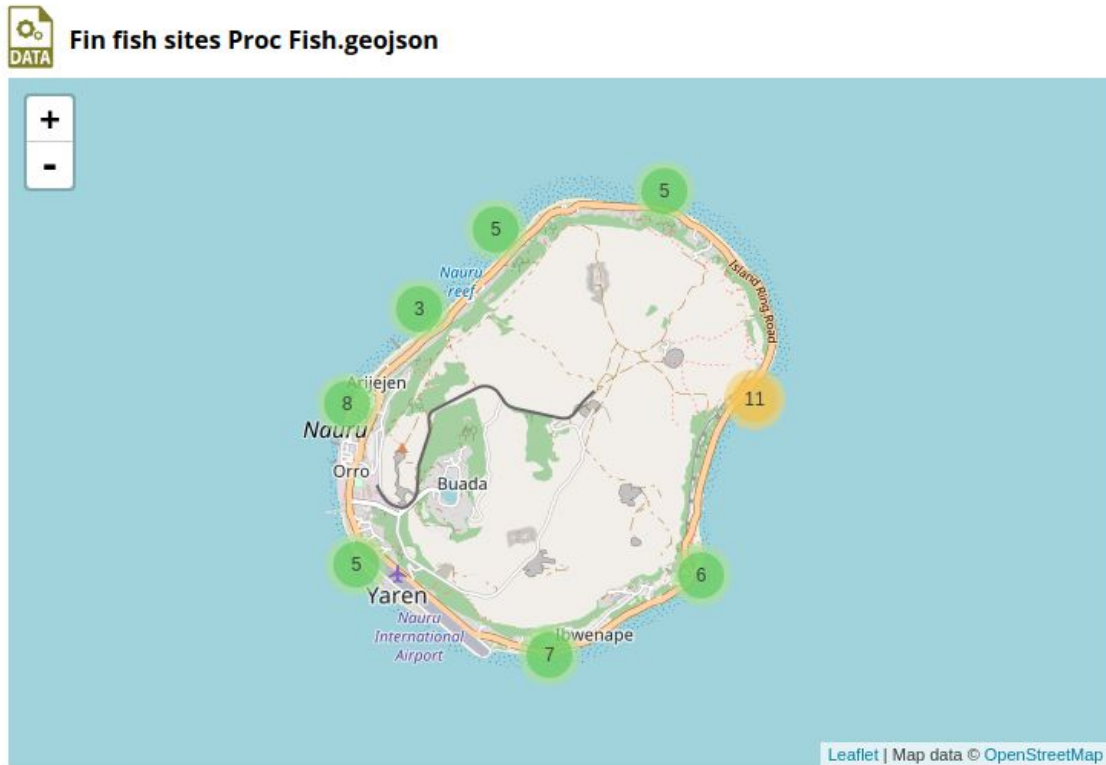
[Import](#) [Cancel](#)

To make CSV data available via the DKAN API, import it into the datastore by going to *Manage Datastore* and clicking *Import*.

GeoJSON Format

GeoJSON is a common format used for spatial datasets and can be generated using GIS software such as ArcGIS and QGIS. DKAN is capable of reading GeoJSON files in order to present spatial datasets directly onto a map. Ensure that *geojson* is selected in the format field when providing a GeoJSON resource.

Procfish sample sites



DKAN can provide map visualisations from GeoJSON files, or CSV file that contain latitude and longitude values.

Step 3: Additional metadata

This step allows for more metadata (information **about** the dataset) to be added for a dataset. If the information is available it is a good idea to add plenty of metadata. All the metadata fields help users find datasets and determine if a dataset is useful to them.

The additional metadata fields can be populated by clicking *Next: Additional Info* after creating a resource record, or by selecting *Edit* from a dataset page.

Additional metadata fields:

Author

The author of the dataset.

Spatial / Geographical Coverage Area

The area that the dataset relates to. Drop a pin or draw a polygon on the map to specify the area

Spatial / Geographical Coverage Location

The name of the area that the dataset relates to.

Frequency

How often a dataset is published.

Temporal Coverage

The start and end date range encompassing the dataset.

Contact Name

The name of the person who can be contacted regarding the dataset.

Contact Email

The email address of the person who can be contacted regarding the dataset.

Homepage URL

The URL of the homepage where more information about the dataset can be found.

Data Standard

The name of the data standard (if any) that the dataset conforms to.

Language

The language used in the dataset.

Related Content

Any other web resources related to the dataset.

Advice for Working With CSV Files

The format of your data makes a big difference to how it can be viewed and used.

Some data formats are *machine-readable*, allowing the data portal to process the files, so they are easier to find and can be published as graphs and charts.

You can make your data *machine-readable* by uploading it in the CSV format.

Formatting data as CSV

Most data-crunching software allows you to simply go `File >> Save As` and choose the CSV format.

Each CSV file is like a single sheet in a spreadsheet document, so you may need more than one CSV file to hold all your data.

Handy checklist

- If you have many 'tables' of data on one sheet, copy each table to its own spreadsheet page and export each sheet as a CSV. That will be a better format for your spreadsheet too.
- In a CSV file, each record must be contained on a single row or column.
- Ensure there are clear and succinct headings for all the data. You can have headings for the columns and for rows, using the first cell at the top of left in each. It is not necessary to have both sets of headings, but it is usually necessary to have one.
- Merged cells are not possible in CSV:
- Flatten your headings into a single cells in which each cell fully describes the content of that column or row.
- Split any merged cells in your data, otherwise the rows will not line up properly.
- If one record spans more than one row, you will need to flatten out the values into columns of a single row.

- If the data contains dates and times or specific units, ensure they are presented consistently. Provide another CSV (data dictionary) which describes the type of data present, its specific formatting and meaning.
- Formulas should be rendered out to their calculated values.
- It is good to include your spreadsheet or database files as additional resources for others who may want to manipulate the data in that format.
- Making CSV files from relational data is a bit more nuanced. You can export results from a query or report as a CSV. Focus on creating a few CSV files which highlight the important aspects of your data.