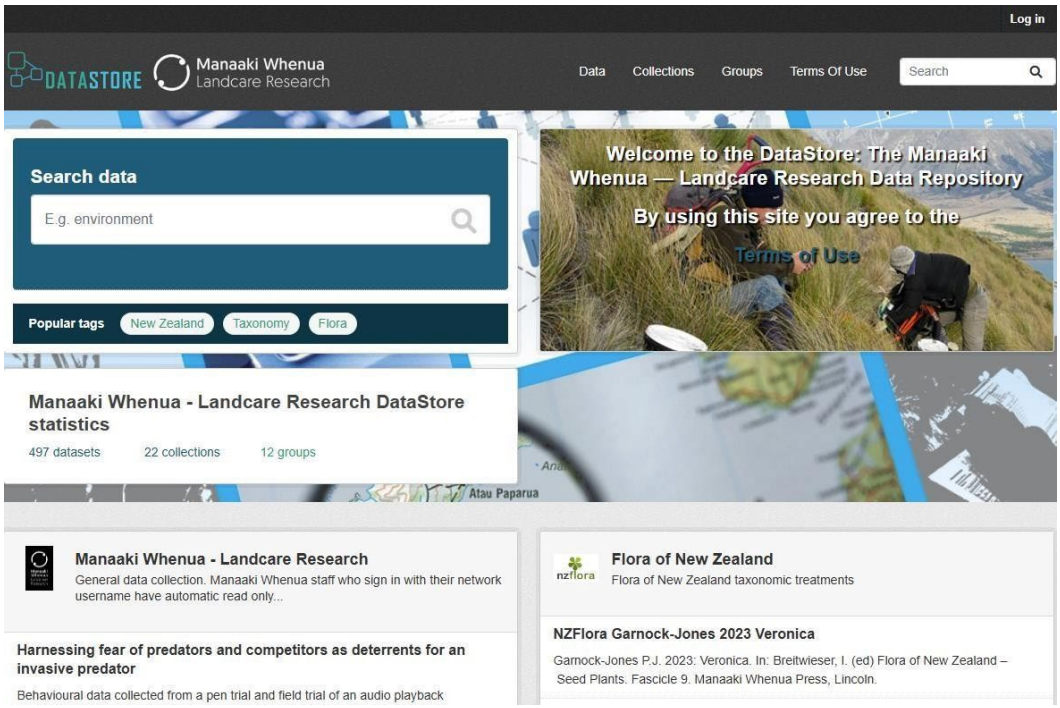
A Guide to Using DataStore: MWLR's Data Repository

Alla Landreth (DataStore Administrator)



Guide for Datastore CKAN v. 2.10.4 (November 2025)

Content

Purpose of this document

This document provides guidance and instructions for effectively using the MWLR data repository DataStore, an open-source platform designed to facilitate the management, discovery and sharing of datasets. The repository aims to maximise the use of MWLR research data and facilitate collaboration by offering a centralised space for staff to upload, access, and reuse high-quality data. Whether you're a dataset contributor or a data user, this document will guide you through the key features of DataStore, from uploading datasets to managing access and ensuring discoverability.

If you have any feedback on this document or DataStore, please send it to the DataStore Administrator.

Approved for release by:

Name: David Medyckyj-Scott, Head of Data Management

Date: 4th December 2024

Updated to reflect the new DataPusher and Stats extensions release dated 10 Dec 2025

Introduction to DataStore

DataStore is Manaaki Whenua - Landcare Research's data repository for data that does not have a natural home in one of our Nationally Significant Databases or where there is no policy or contractual requirement for the data to be deposited in another data repository, e.g., one associated with a journal.

It is primarily for the purpose of storing and sharing data collected or created from publicly funded research (as required by the Government), however if you have written permission from your client, you may deposit other datasets.

Some government agencies (both central and local) may be agreeable to you depositing and sharing data from projects they have funded, as they are also subject to government requirements around open data and data sharing.

Anyone can browse the DataStore, but to get further rights you need to be Manaaki Whenua affiliated user.

Warning - Datasets flagged as *publicly accessible* in DataStore will show up in Google search results and other internet searches. This increases the visibility of your data but may result in enquiries from researchers and other individuals who may then require assistance.

Prerequisites for depositing datasets

DataStore has been made a relatively open system. MWLR staff have the ability to publish data directly; you do not currently[1] need someone to login into DataStore to approve publication of your data. With that ability comes certain responsibilities and trust. Please read the DataStore Terms of Use carefully.

Before depositing and making public any data, you must:

- confirm there are no contractual requirements on where and how data should be shared. The contract may also state what metadata is required and what other information is necessary information for onward users to appropriately understand the dataset and any limitations. For example, MBIEs Open Research Policy requires data to be shared with other researchers', to be deposited in subject data repositories or made available through online institutional repositories like DataStore.

[1] This may change in the future by the implementation of a data publication pipeline similar to paper and report publication.

- consider whether DataStore is the appropriate repository for the type of data you want to share. For example, while DataStore may be appropriate for simple geospatial data, the LRIS Portal may be more suitable for more typical or complex geospatial data[2].
- confirm MWLR owns the data or get written permission from clients to deposit data into DataStore, and, potentially, make the data publicly available. Make sure you retain any communications from a client regarding them agreeing for data to be deposited in DataStore and, where applicable, released.
- consider if there are any confidential/privacy/sensitive data/ethical issues that need to be addressed before depositing the data and making the data public. This may require the data to be anonymised, generalised or deidentification of people's names or locational information to create a shareable version.
- consider whether the data is associated with a publication that has gone through a peer review process (internal and/or by an external journal)?
    - If Yes: we are accepting this as proxy that if the paper/report is signed off as of acceptable quality, the underlying data should also be.
    - If No: you should get approval from your PL (or RPAL if the PL has delegated) that they are happy the data is of appropriate quality to deposit ***and*** make public.
- consider under what licence others may use the data. Do not make any dataset public without specifying an appropriate licence.

To upload Māori data into DataStore, approval is typically required from the Māori community or iwi who own, have custodianship over or an interest in the data. Adherence to data sovereignty principles would also be important considerations.

You should read MWLR's Data Management Policy to familiarise yourself with the general requirements around data management.

If you are unsure of anything, consult the DataStore Administrator.

Metadata Requirements

Each dataset in DataStore has a metadata record - information about the data. For example, the title, date of the data, tags (keywords), what formats it is available in, what licence the data is released under, contact details, etc.

[2] The LRIS Portal offers similar features to DataStore but can handle larger spatial datasets and offer users more customisable ways to download the data (e.g., extent, file format, projection). Your data would also sit alongside other MWLR spatial data in a system known by many spatial data users and thus get more visibility.

You need to provide sufficient metadata to make the data understandable and usable. In some cases, this may require you to provide additional supporting material.

DataStore requests you to complete a few basic metadata fields, but more detail as to the specifics of the data, such as field definitions and/or methodology, can be deposited alongside the data. None of this is difficult, and there are tools to help in some instances.

More detailed metadata can also be loaded as a Resource (explained below), e.g., descriptions of what fields in a spreadsheet mean, the units of measurement used, the properties, data collection methods, how the data was processed, what quality control was undertaken, provenance of any data used to create the data in the dataset, etc.

Login Instructions

MWLR staff use their MWLR network login credentials to login into DataStore.

Important: Your username is the first part of your email address, **not** your full email address.

Please contact the DataStore Administrator if you require non-MWLR staff to have access.

Key DataStore concepts it is important to understand

To use DataStore effectively you need to understand some key concepts about how data is organised.

Organisation

Manaaki Whenua - Landcare Research is the recognised organisation for DataStore. The organisational administrator has control over the MWLR collections, groups, datasets and members.

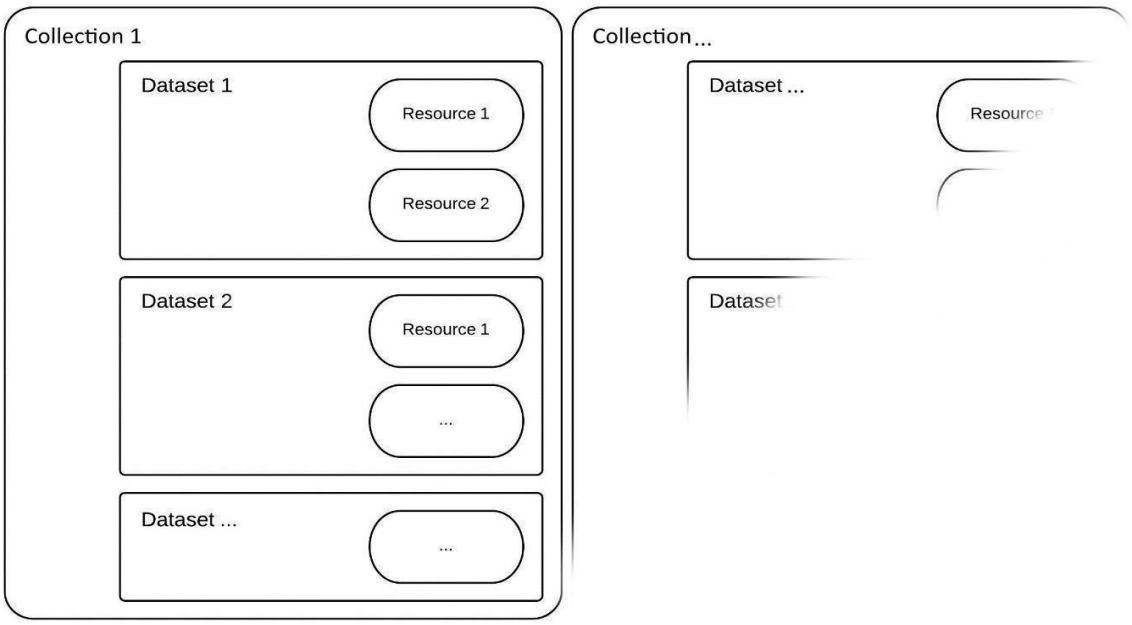
Collections

Collections are just that – Collections of various datasets and are the primary way to control who can see, create and update datasets in DataStore.

Datasets

Datasets are ‘packages’ of data created under a collection, and comprise two things:

- an administrative metadata record
- one or more *Resources* (essentially files or data objects that are associated with a **Dataset)** which hold the data itself (an uploaded file, or link to a file hosted elsewhere).



Datasets effectively package together related information in one place for easy access, e.g., your dataset could contain your research data, an explanation of the data fields (metadata), the associated project's data management plan (DMP), and a copy of (or link to) a journal paper or report the results of which were based on the data you are depositing. Including a copy or link to the latter is an easy way to provide further detail on the research methodology.

Important: Currently, new Collections must be setup by the DataStore Administrator who will then assign a user as a Collection administrator. This person can add members to the Collection and assign them rights.

Collections can be setup around a range of use cases, e.g.,

- a science discipline
- a large/significant project
- a Portfolio or Science Team
- a laboratory or research group
- a significant client who has given written permission for us to hold (and potentially share) data from research they have funded.
- some other logical way of arranging datasets…

When thinking about where to put your data, look at what Collections have already been setup, and whether the data needs to have restricted access. Data that is to be public can be placed into a general collection. However, private datasets may need to be placed in a collection with a more targeted membership. In future we hope to be able to add further controls over access permissions.

Groups

Groups are another way public[3] datasets can be arranged together. Groups allow the organisation of data by themes, topics, or other logical categories, making it easier for users of DataStore to find and access related datasets. A dataset can belong to multiple Groups (but only a single Collection). Users who can see and have access to Groups do not have control over authorisation or access rights; this is controlled by the Collection administrator.

If you would like a new Group established, contact the DataStore Administrator[4]. You will be assigned administration rights for the new Group and can then add other people to the Group so they can upload datasets that they work on. Anyone can see the (public) datasets within a Group, but only Group members can add or remove datasets to/from the Group.

If you see a Group you would like to join, click on Group, and then on the 'read more' link under the Group description at the left-hand side to see if there is contact information for the Group administrator. Alternatively, contact the DataStore Administrator.

User rights

Each dataset can belong to a single Collection, and each Collection controls access to its datasets.

A user who has not logged in or an unregistered user will have minimal *Visitor* rights. Visitors can view and download datasets that are marked as public.

Registered users can be assigned one of 3 roles: *Member*; *Editor*; *Admin* for specific Collections. You can have different roles for different Collections.

To upload (or edit) a dataset you need to be a registered user and be granted *Editor* or *Admin* rights to a Collection.

Access rights

Datasets visibility can be marked *Private* or *Public*:

- Private datasets are only visible to users who are members of the Collection the dataset belongs to. Private datasets do not show up in searches (for anyone).
- Public datasets can be discovered and seen by anyone who visits DataStore and will be discoverable by Google and other search engines.

[3] It is not recommended to add private datasets to groups.

[4] DataStore was originally configured so that any registered user could create their own groups. However, the site was spammed by bogus users registering and creating 'fake' groups – thus we have had to limit establishment of groups to the DataStore Administrator.

Appendix 1 provides more details on what permissions different users can have.

Have a go

DataStore is built on the open-source software CKAN. If you want to practice/explore the functionality of CKAN, visit the demo site https://demo.ckan.org (but don't upload real data of course!). Note: the demo site has some settings that differ from our installation, but you will get a good feel for how things work.

Collection Management

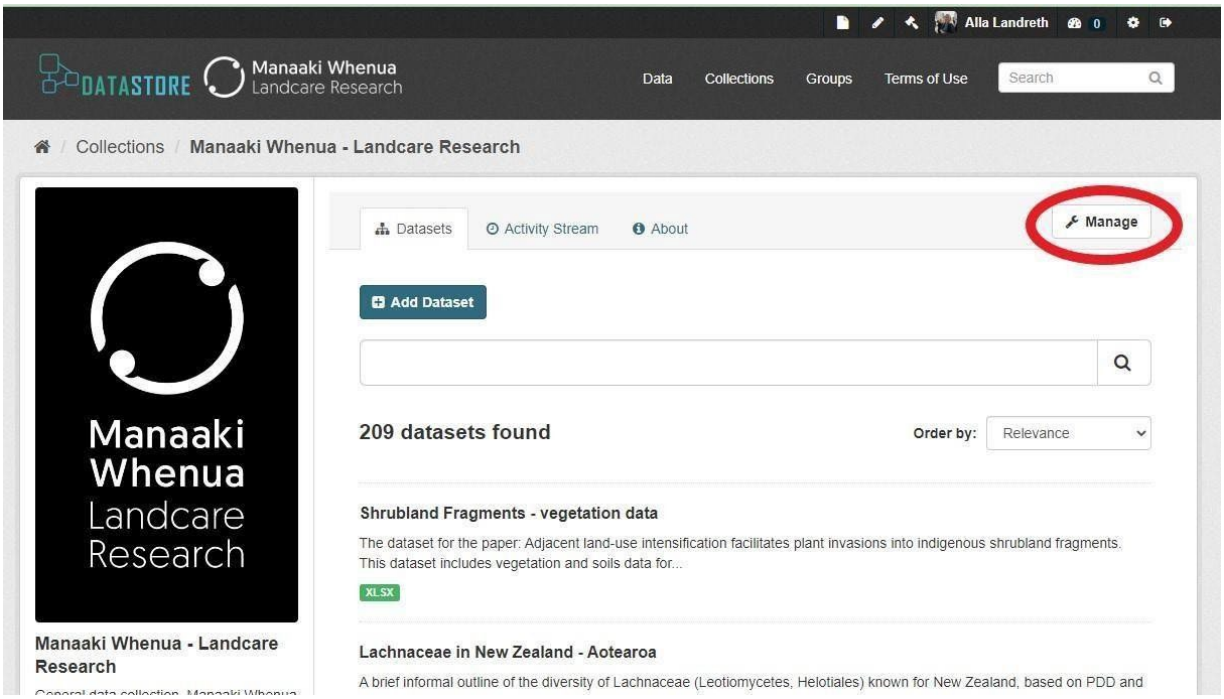
Creating a Collection

The DataStore Administrator is the only person who can create a new Collection in DataStore – contact them if you think you need a new Collection.

You will need to provide a name and short description for the Collection and, optionally, an image (provide a file or a link to the image) for the Collection's home page.

Managing a Collection

When the DataStore Administrator creates a Collection, you (or a person you specify) will be made administrator ("Admin") of the Collection. On the Collection's homepage Admins see a "Manage" button.



Clicking “Manage” displays the Collection admin page. This page has three tabs:

- Edit: Here you can edit the information supplied when the Collection was created (title, description, and image).
- Datasets: Lists the datasets belonging to the Collection. You can switch them to have either private or public status or delete a dataset.
- Members: Here you can add and remove users and change their access roles in the Collection. You will need to know their name or username.

Datasets and Resources

Adding a new dataset

If you are an Editor or Admin of a Collection, you can add or edit datasets.

Step 1. You can access DataStore's "Create dataset" screen in three ways.

a) Select the 'Data' link at the top of any page, then click the "Add Dataset" button.

b) Select the "Collections" link at the top of a page. Now select the page of the Collection to want the dataset to be added to. Then click the "Add Dataset" button from the Collections Datasets tab.

c) Click on your username (top right) to access your dashboard. Click on the “My Datasets” tab, then click the "Add Dataset" button.

Step 2. DataStore will ask for the following information about your data. (The actual data will be added in step 3.)

- *Title* - this title will be unique across DataStore, make it brief but specific and informative, e.g., " Upper Taiari Scroll Plain - Wetland Vegetation Survey 2023 " is better than "Vegetation survey".
- *Description* - You can add a longer description of the dataset here, including information such as where the data is from and any information that people will need to know when using the data.
- *Tags* - here you can add tags (keywords) that will help people find the data. Examples could be "population dynamics", "soil carbon", "possum".

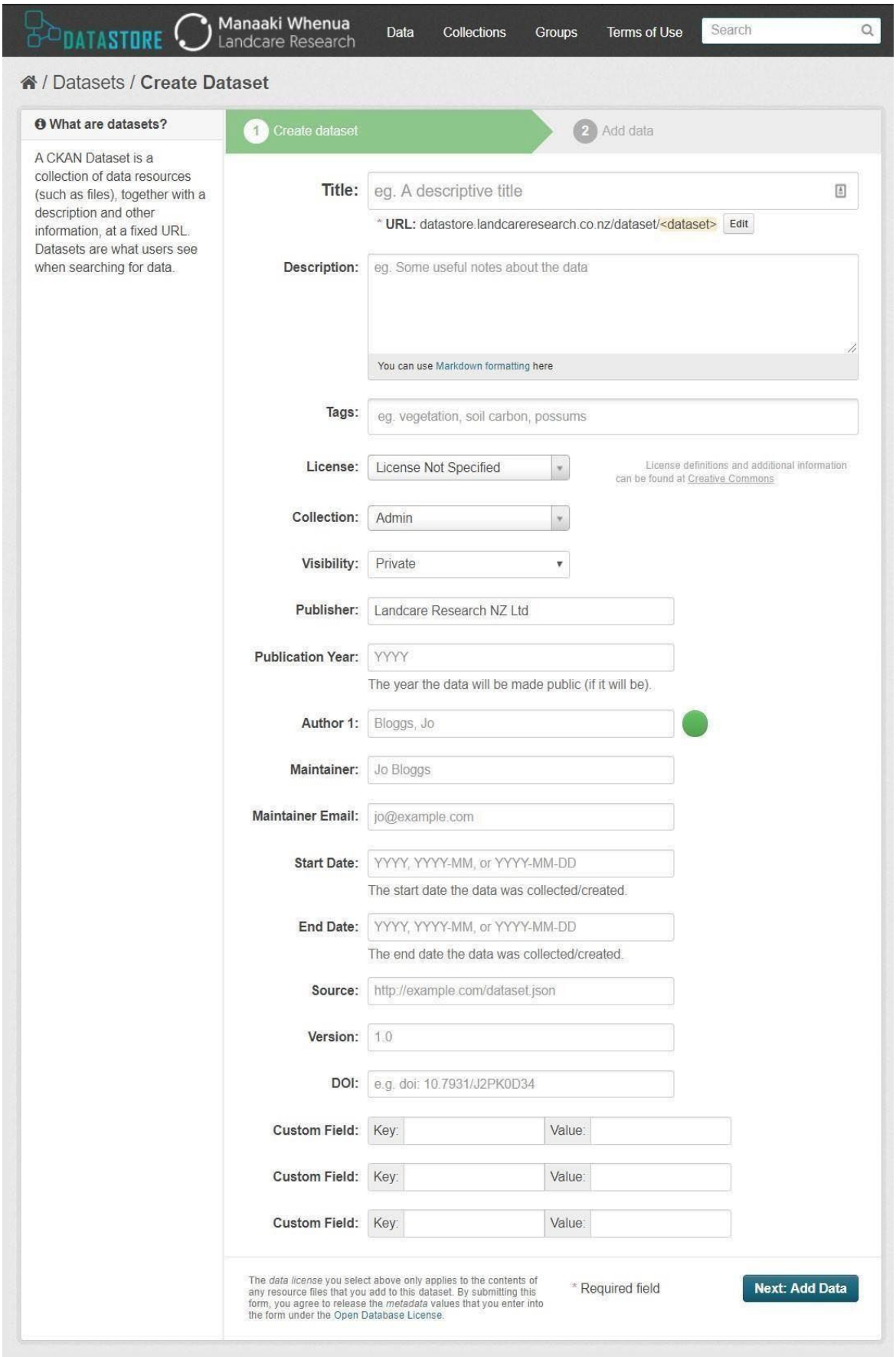
  As you start to type a tag previously used tags will appear as suggestions. Select a required tag from the list, or, if the tag you want is not suggested, continue to type in the word(s) you want to add as a tag. Press the <Enter> key to add the tag. If you enter a tag wrongly, you can use its delete (x) button to remove it before saving the dataset.
- *License* - it is **critical** to include license information so that people know how they can use the data. A number of options are available from the drop-down menu, e.g., CC-BY 3.0 NZ (Attribution), Some research funders will state what licence should be used if data is published, e.g., MBIE encourages the use of the Creative Commons Attribution licence (CC-BY), permitting free use and reuse subject to proper attribution.
- *Collection* – from the drop-down menu choose the one the dataset belongs in. Ensure the default chosen is the correct one before you proceed.

- *Visibility*: It is very important the 'Visibility' is set correctly. A **Public** dataset can be seen by any visitor to the site. A **Private** dataset can only be seen by members of the Collection owning the dataset and will not show up in searches by any users.
- *Publisher* – The default is 'Landcare Research NZ Ltd' (generally there is no need to change this).
- *Publication Year* - The year the data will be made public (if it will be).
- *Author(s)* - The name(s) of the person(s) responsible for producing the data. Use the green button to add additional authors.
- *Maintainer* / *Maintainer e-mail* - details of the person who is the maintainer of the data, i.e., the best contact person for questions about the data. Remember to update this if the person leaves MWLR.
- *Start Date / End Date* - The start and end dates for the period the data relates to, e.g., when it was collected/created. This can be in written as YYYY, YYYY-MM, or YYYY-MM-DD format.
- *Source* – If the data is derived from other datasets, state that here, e.g., if the data was originally sourced from NVS you could provide a link to NVS.
- *Version* - if more than one version of the data set will be deposited in DataStore, a version numbering system should be used, e.g., 1.2. Alternatively, if the data is regularly collected and published you may enter the date of a particular survey, e.g., 2008, or October 2013.
- *DOI* – If the data has been issued a DOI (recommended for all publicly available datasets) it will be recorded here. Usually, the DataStore Administrator will add this when a MWLR DOI is issued, or you can add one if the DOI was created for you by another party.
- *Custom fields* - If you want the dataset to have another field, you can add the field name and value here. Examples might be "Funder" or "Job Code". The number of custom fields available will increase as you use one.
- A special custom field 'spatial' can be used to define and display the location of the data and thus enable spatial searching/filtering of datasets. See Appendix 2 for more information about this.

Important: The only mandatory fields on this page are the Title and Collection. However, the Core metadata fields MWLR regard as best practice to include are

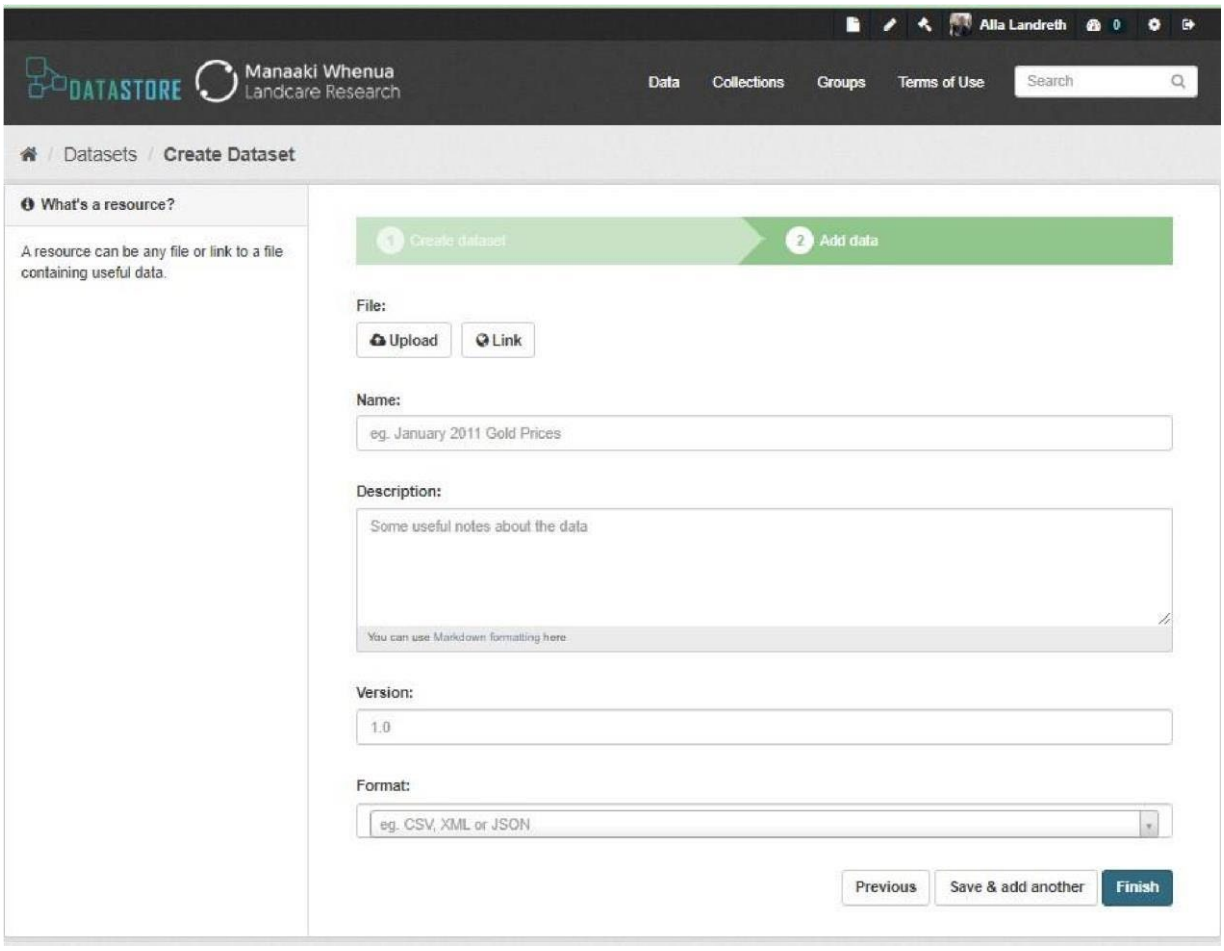
- title
- short description,
- the license information (**do not** make a dataset public without specifying an appropriate license)

- start date / end date
- publisher
- author(s) and
- maintainer
- keywords
- DOI (if dataset will be made public)



Dataset creation screen

Step 3. When you have filled in the information on this page, click the "Next: Add Data" button. DataStore will display the "Add data" screen.



This is where you will add one or more "Resources". Obviously, the key Resource is the one for the data for this dataset.

- If the data to be added is in a file on your computer/the network, select "Upload". DataStore will open a file browse dialog box for you select the file.

  Warning: DataStore has an upload size limit of 50 GB for resources.

- To link to a file held outside the DataStore select "Link" and enter the location, e.g., "http://example.com/mydata.csv"

Add the other information asked for on the page. DataStore does not enforce this, but it is good practice to add it to ensure a user is clear of the content of each Resource.

- *Name* - a name for this Resource, e.g., "Vegetation survey 2014, CSV". Different Resources in the dataset should have different names.
- *Description* - a short description of the Resource.

- *Format*[5] - the file format of the Resource, e.g., csv (comma-separated value 'spreadsheet'), tsv (tab separated value 'spreadsheet'), xls or xlsx, pdf, txt, jpg, html, R (R script file), py (python script), nex (nexus file) … [don't put the 'dot' in the format]

DataStore will try to provide users with previews of the data where possible. For this to happen, it is important to use the correct extension on your files, and in this field.

If you have additional Resources (files or links) to add to the dataset, select the "Save & add another" button and repeat step 3.

Step 4. When you have finished adding Resources, click "Finish". Note – you can add additional Resources later (see Editing a dataset below).

DataStore creates a dataset record and shows you the result.

You have finished!

If you set the visibility of the dataset as Public, you should now be able to find your dataset by typing the title or some relevant words from the description into the search box on any DataStore page. We will say more about searching for data in the section "Finding data".

Tabular data and Previews

For tabular data, particularly tsv and csv files, as well as storing a copy of the file for download, DataStore imports the data into a central database. This data can be previewed by users and preliminary data investigations done (e.g., simple graphs, displaying spatial data (with Lat/Longs) on a map). You can refine the imported data using the "Data Dictionary" tab of the Resource edit screen. Here you can override the type of data, modify labels to make the field name more meaning full, and add descriptive information about a data field.

Tabular data stored in the database can also be accessed using the CKAN-R library.

Editing a dataset

You can edit the dataset you have created, or any dataset owned by a Collection that you are an *Editor* or *Admin* member of.

Do not attempt to edit datasets you are not responsible for – DataStore tracks who makes changes!

Be careful if you are making changes, there can be flow on implications in some cases, e.g., if you are considering changing the license (other than from 'not specified'), or if the dataset has been issued a DOI – consult the DataStore Administrator for advice.

[5] Any file format can be stored in DataStore. Some formats are more robust for long term storage e.g., tab or comma separated value text files are preferable to excel (recommend store as tsv/csv, or store both formats).

Step 1. Go to the Dataset's page, click the "Manage" button (upper right).

Step 2. DataStore displays the "Edit metadata" screen.

You can edit any of the fields (Title, Description, etc.), change the visibility (Private/Public), and add or delete tags or custom fields. For details about these fields, see "Adding a new dataset" above.

Step 3. When you have finished, select the "Update dataset" button to save your changes.

Important: being granted editor/admin rights comes with responsibility – you should only edit datasets that you are responsible for as maintainer, or an author, and those who need to know should be made aware you are making changes.

It is ***very important*** that data that has been issued with a DOI is not changed.

If there is an issue with a DOI discuss this with the DataStore Administrator.

If you correct an error in the data, make a note of this in the metadata associated with the dataset. If needs be, add a document documenting the error and the fix as a Resource.

Embargoed Data and Placeholder PDFs

In cases where a dataset is under embargo and cannot yet be made publicly available, DataStore supports the use of a downloadable embargo notice PDF as a temporary placeholder resource. This allows the dataset's metadata to remain visible for discovery and citation while clearly informing users that access to the data itself is restricted.

How to Apply an Embargo Placeholder:

1. Upload a Resource named 'Embargo Notice' or similar.

2. Use the prepared embargo notice file (e.g., Dataset Under Embargo.pdf), which includes:

- A short explanation that the data is under embargo.
- The expected embargo release date.
- Contact details for the dataset maintainer (e.g., project lead or data steward).

3. Set the resource format to PDF.

4. Add a clear description, e.g.:

"This dataset is currently under embargo. Please refer to the attached notice for more information, including the release date and contact person."

This approach allows DataStore users to discover embargoed datasets while respecting access restrictions and ensuring transparency.

A standard embargo notice template is available from the **DataStore Administrator – Alla Landreth** or can be downloaded from the "About" page.

Important: being granted editor/admin rights comes with responsibility – you should only edit datasets that you are responsible for as maintainer, or an author, and those who need to know should be made aware you are making changes.

It is very important that data that has been issued with a DOI is not changed.

If there is an issue with a DOI discuss this with the **DataStore Administrator**.

If you correct an error in the data, make a note of this in the metadata associated with the dataset. If needs be, add a document documenting the error and the fix as a Resource.

Adding, deleting and editing Resources

Step 1. Go to the Dataset's page, click the "Manage" button (upper right).

Step 2. Click on the "Resources" Tab. Click on an existing Resource to edit or delete it or click "Add new Resource" to add a further Resource to the dataset.

Step 3. You can edit the information about the Resource or change the linked or uploaded file. For details on this, see step 3 of "Adding a new dataset", above.

Step 4. When you have finished editing, select the button marked "Update Resource" (or "Add", for a new Resource) to save your changes. Alternatively, to delete the Resource, select the "Delete Resource" button.

Be careful deleting a Resource – this cannot be undone.

Deleting a dataset

Step 1. Go to the dataset's "Edit dataset" page (see "Editing a dataset", above).

Step 2. Select the "Delete" button.

Step 3. DataStore displays a confirmation dialog box. To complete deletion of the dataset, select "Confirm".

Warning: The "Deleted" dataset is not completely deleted. It is hidden so it does not show up in any searches, etc. However, by visiting the URL for the dataset's page, it can still be seen (by users with appropriate authorisation), and 'undeleted' if necessary. This may not be what you want. If it is important to completely delete the dataset, contact the DataStore Administrator.

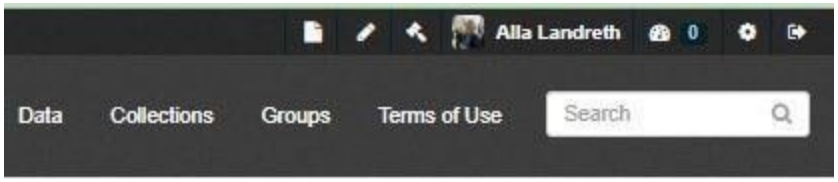
Periodically deleted datasets will be erased and will no longer be recoverable. If you accidentally delete a dataset, contact the DataStore Administrator asap.

Again, a reminder that if you delete a *Resource* from within a dataset, it cannot be recovered.

Important: Do not delete a dataset or Resource for which a DOI has been issued without consulting the DataStore Administrator.

Finding data

Only public datasets are returned in general searches of DataStore, no matter whether you have permission to see private datasets or not. Searching within a Collection will show private datasets if you are a member of that Collection.



Searching the site

To find datasets in the DataStore, type a search term into the search box on top right of any page.

DataStore displays the first page of results of your search. You can

- view more pages of results
- repeat the search, altering or adding terms
- restrict the search to datasets with particular tags, data formats, etc using the filters in the left-hand column

Search filters

If there are many results, the filters can be a very helpful way to reduce the number of results. You can combine filters, selectively adding and removing them, and modify and repeat the search with existing filters still in place.

The filter options appear at the left-hand side of the browser. As you apply a filter the results will be restricted to datasets matching the criteria selected so far.

You can filter by: *Collection*; *Group*; *Tag* (keywords); file *Format*; *License*; and *Author.*

Tip – to see all the (public*[6]) datasets you are an author/co-author on, click “Data” from the top menu, then find your name in the list of author filters. This will provide a view of datasets including those that you have not uploaded yourself but are a co-author of (which are visible in the ‘My Datasets’ view). To get the most from this feature, ensure your name is entered in a consistent way when you are listed as an author.

Geographic (spatial) searching…

For datasets that are spatially tagged (see Appendix 2) it is possible to search for datasets by selecting an area on a map.

Step 1. Click on the “Data” link at the top of any page.

Step 2. Click the ‘pencil’ button. Drag around the region you wish to find data within. If you need to zoom in, either:

- Draw roughly to begin, click ‘Apply’, then refine the search.
- Or, click (not drag) on the map, the cursor changes to a ‘hand’.

  Then use the +/- buttons to zoom the map and click and drag to reposition the view to your area of interest. Then click the ‘pencil’ button and drag to define the area of interest.

Step 3. Click ‘Apply’ to search for datasets with a spatial footprint within/overlapping the area you defined. You can then apply further search terms and/or filters to refine the search.

Searching within a Collection

If you want to look for data owned by a particular Collection, you can search within that Collection from its home page in DataStore.

Step 1. Select the "Collections" link at the top of any page.

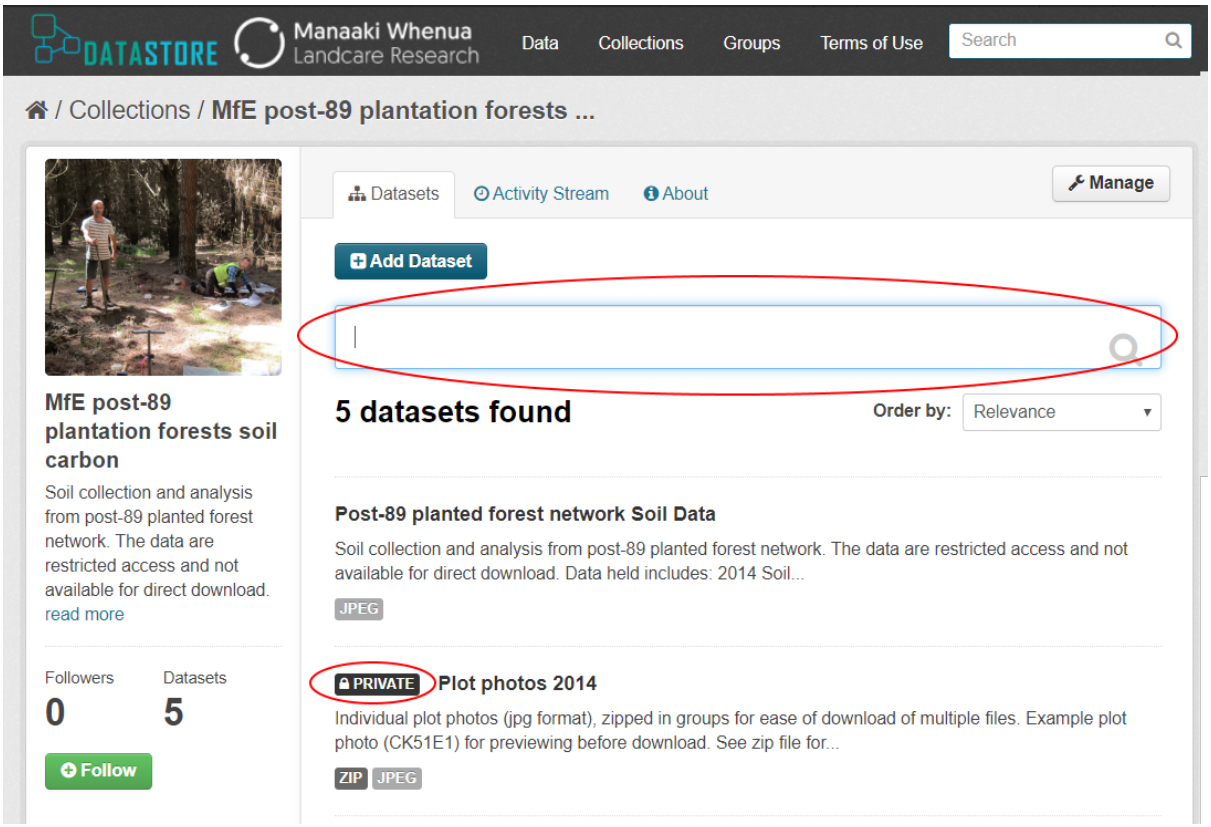
Step 2. Select the Collection you are interested in. DataStore will display the Collection's home page.

Step 3. Type your search in the main search box on the page.

[6] Private datasets will be included in the results if you apply the filter within a Collection.

The search will return results restricted to datasets from the Collection. You can again apply filters to the search results to further refine the search result set.

If you have member and have access to the Collection, private datasets will also be included in your search results.



DataPusher

DataPusher is a service used to **automatically** load uploaded files into the DataStore for indexing and preview.

The CKAN DataStore is the database that allows you to search, filter, query, and preview datasets. DataPusher acts as a bridge, taking a file (such as a CSV, TSV or Excel spreadsheet) and converting it into structured database records. It supports Excel files as well as CSVs, and once data is loaded intoDataStore, it becomes immediately available to:

- Explore through the DataStore web interface
- Download in different formats

- Query directly via the DataStore API

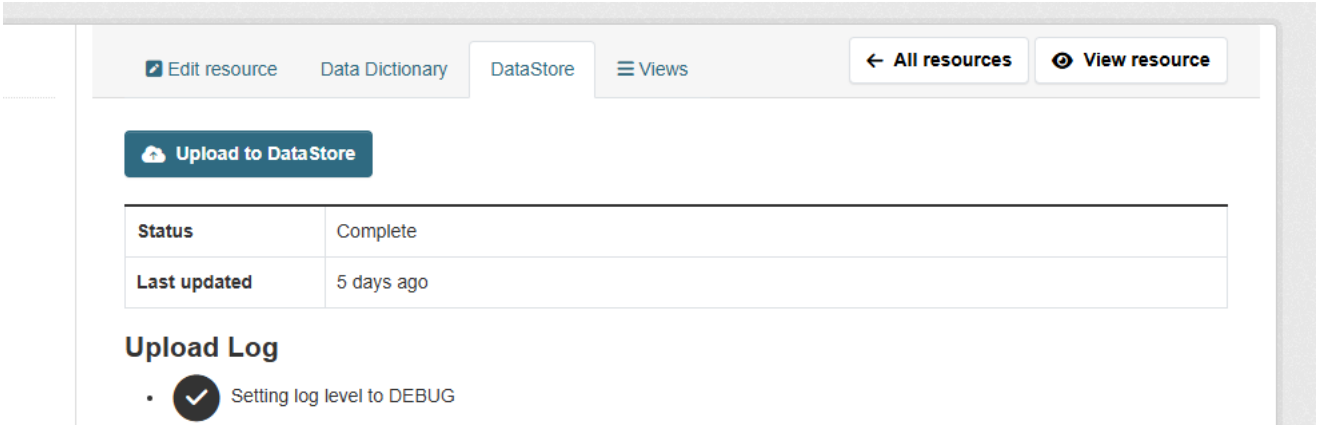
Tips for Best Results

1. **Make sure your data are well organised**:
   - Use clean CSV or XLSX files: Avoid formulas, comments, colour formatting, or hidden rows/columns.
   - Include one clear header row: DataPusher uses the first row to create field names.
   - Keep column names simple: No special characters; avoid spaces if possible.
   - Check for merged cells: DataPusher cannot process merged cells.
   - Remove blank rows or columns: Leading/trailing empty rows often cause errors.
   - Use consistent data types per column: Don't mix text and numbers in the same column.
   - Ensure valid dates: Use YYYY-MM-DD or a consistent date format.
   - The current file size limit is 50 GB

How to Upload Data with DataPusher

1. **Go to your dataset.**
   - Click the **"Manage"** to switch into editing mode.
   - Click **Add Resource**.
2. **Upload your file.**
   - Choose a CSV or Excel file (DataPusher supports both).
   - Add a description if needed.
   - Click **Save**.
3. **Wait for processing.**

- DataPusher automatically starts and loads uploaded files into the DataStore for indexing and preview Update your browser to refresh the page to see the latest Data Dictionary, preview, and DataStore table once DataPusher has finished processing.



4. **Check the status.**
   - If successful - the “DataStore active” message will appear, and you can preview the data.
   - If failed - An error message will display (see Troubleshooting DataPusher).
5. **Use your data.**
   - Browse and search in the web interface – Data Explorer or Table views are available

   - Download filtered views.
   - Query the dataset via the DataStore API.

Data Dictionary

The **Data Dictionary** in DataStore describes the structure of your dataset once it has been loaded into the **DataStore**. DataPusher only processes tabular files, such as: CSV, TSV, XLSX (Excel). These are files that can be parsed into rows and columns. It provides information about each column, including:

- Column name
- Data type (text, number, date, etc.)

- Description of what the column means

When DataPusher Plus loads your file, it tries to **guess the column types automatically**. Sometimes these guesses may not be perfect (for example, a column of numbers with mixed characters might be loaded as text). The Data Dictionary feature lets you **review and adjust these settings,** so your dataset is easier to understand and query.

Maintaining a Data Dictionary is especially useful when:

- Sharing data
- Running queries via the API (ensures correct data types).
- Using large or complex datasets

You can edit the Data Dictionary from the dataset's **Manage - Data Dictionary** tab after DataPusher has finished processing. We strongly recommend you review the contents of the Data Dictionary DataPusher has created.

Troubleshooting DataPusher

If DataPusher fails to upload your resource to the DataStore, an error message will be displayed. Below are common issues and how to resolve them:

1. File Format Issues

- **Problem:** Unsupported or corrupted file type.
- **Solution:** Ensure the file is in a supported format (CSV, Excel, TSV). Re-save the file using a standard editor before re-uploading.

2. Invalid Characters or Encoding

- **Problem:** Error messages mentioning "invalid byte sequence" or encoding problems.
- **Solution:** Save the file with **UTF-8 encoding**. Check for hidden or special characters in column headers or data rows.

3. Column Names and Data Dictionary Errors

- **Problem:** Error stating multiple values for a non-multi-valued field (e.g., *Solr error: multiple values encountered for author*).
- **Solution:**
    - Ensure each field (column) contains only one value unless explicitly configured as multi-valued in the schema.
    - Avoid commas or semicolons inside single-value fields unless properly quoted.

- Review your **Data Dictionary** in DataStore to confirm field types and constraints. The Data Dictionary can still be created even when the data itself fails to load. The activity log will show an error explaining why the load did not complete.

4. Large File Size or Timeout

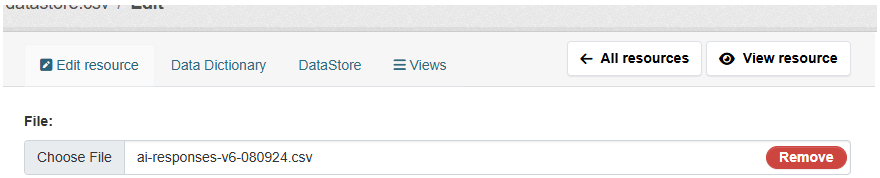
- **Problem:** DataPusher times out or fails to complete for large datasets.
- **Solution:**
  - Split the dataset into smaller files.
  - Consider using the **DataStore API** for direct data uploads.
  - Check server logs for timeout limits. To do this press **F12** (Windows/Linux), Go to the **Network** tab and look at the Status codes. The status code to watch for is 504 Gateway Timeout → the server or proxy timed out.

5. Network or Server Errors

- **Problem:** Error messages like *502 Bad Gateway* or *Connection refused*.
- **Solution:**
  - Retry after a few minutes in case of temporary downtime.
  - Contact the DataStore administrator if the issue persists.

6. General Steps to Resolve Issues

- Re-upload the file after fixing formatting issues. To avoid data duplication, DataStore deletes the current resource before a new version is uploaded.



- Review the displayed error message carefully—it usually points to the problematic field.
- If problems persist, raise a support ticket with the DataStore administrator, including the error message and dataset link.

Exploring datasets

When you have found a dataset, you are interested in and selected it, DataStore will display the dataset page.

The dataset page displays

- the name, description, and other information about the dataset
- links to and brief descriptions of each of the Resources

The Resource descriptions link to a dedicated page for each Resource.

Each Resource page includes information about the Resource and enables it to be downloaded.

Many types of Resource can also be previewed directly on the Resource page. Tabular data ‘spreadsheets’ in comma or tab separated value text files (CSV, TSV) are previewed in a grid view, with map and graph views also available if the data is a suitable form. An Excel (XLS, XLSX) format can be added to a Grid view, but by default, it will use an Office docs previewer to display the data (this viewer only works if the data is public). The Resource page will also preview Resources if they are common image types (e.g., jpg), PDF, or HTML.

The dataset page has two other tabs:

- Groups – where you can add a dataset to a Group(s) which you have been granted permission to.
- Activity stream – shows the history of recent changes to the dataset.

“Following” a dataset

If the dataset is of interest to you going forward, you can opt to be notified of changes made to it by using the "Follow" button on the dataset page. There is more about this feature in the section "Managing your news feed" below. You must have a user account and be logged in DataStore to use this feature.

Map view

Clicking on a Resource will (depending on the file format) display a preview of the data table, and you can then display spatial data on a map or construct various graphs.

Data with geographic information (e.g., WGS84 latitude/longitude) can be displayed on a map

Personalisation

DataStore provides features to personalise the experience of both searching for and publishing data. You must be logged in to use these features.

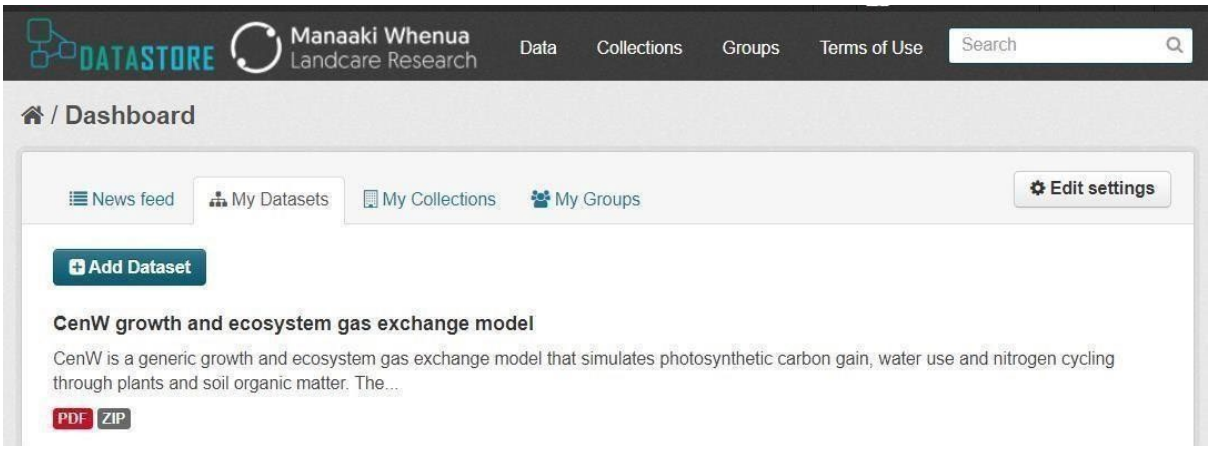
Managing your News Feed

At the top of any page, select the dashboard symbol  (next to your name). DataStore displays your News Feed. This shows changes to datasets that you follow, and any changed or new datasets in Collections that you follow. The number by the dashboard symbol shows the number of new notifications in your News Feed since you last looked at it.

As well as datasets and Collections, it is possible to follow individual users to be notified of changes that they make to datasets. If you want to stop following a dataset (or Collection or user), go to the dataset's page (e.g., by selecting a link to it in your News Feed) and select the "Unfollow" button.

My Datasets, Collections, and Groups

From the dashboard you can also see the datasets you have deposited or Collections and groups that you have administer.



Managing your user profile

You can change the information about that DataStore holds about you, including what other users see about you, by editing your user profile. Click the username symbol at the top of any page.

DataStore displays the user settings page. Here you can change:

- About (add information you want other users to know about you such as your ORCID id, areas of research, etc.
- A profile Picture

Make the changes you require and then select the "Update Profile" button.

Statistics

https://datastore.landcareresearch.co.nz/stats

Understanding the Dashboard:

The DataStore Stats Extension provides insights into various aspects of DataStore, including:

- **Total number of datasets:** The overall count of datasets published on the site.
- **Dataset revisions per week:** A measure of how actively datasets are being updated.
- **Top-rated datasets:** Datasets that have received the highest ratings from users (if rating is enabled).
- **Most-edited datasets:** Datasets that have undergone the most revisions.
- **Largest groups:** Groups with the highest number of associated datasets or members.
- **Top tags:** The most frequently used tags across all datasets.
- **Users owning most datasets:** Identifies users who have published the largest number of datasets.

Dataset-Level Statistics

Each dataset includes a **Usage / Statistics** section showing detailed activity metrics.

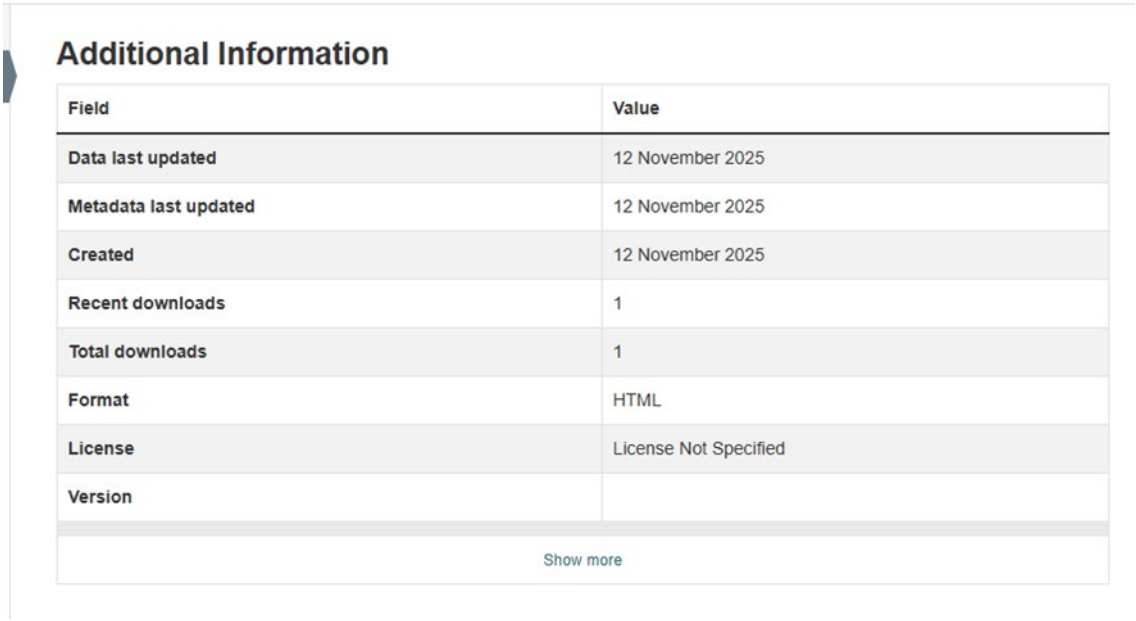
What You Can See:

Recent dataset views – number of views in the recent past (last 7).

Total dataset views – cumulative views since the dataset was published**.**

Resource-Level Statistics

Each resource includes a **Usage / Statistics** section showing detailed activity metrics.



Additional Information

Field	Value
Data last updated	12 November 2025
Metadata last updated	12 November 2025
Created	12 November 2025
Recent downloads	1
Total downloads	1
Format	HTML
License	License Not Specified
Version	

Show more

What You Can See:

Recent resource downloads – how many times each resource was downloaded in the last 7 days

Total resource downloads – lifetime number of downloads for each individual resource.

These metrics help dataset owners to understand:

- Short-term engagement (recent views/downloads)
- Long-term impact (total numbers)
- Which datasets and resources are performing best
- Which datasets attract the most attention
- Which specific resources users find most valuable
- Usage growth over time

Appendix 1 - Authorisation and associated permissions

Rights/Permissions	Visitor (not logged in)	Authenticated User	Collection Editor	Collection Admin	SysAdmin
Has user profile	NA	Y	Y	Y	Y
Able to be added to Collection or Group	NA	Y	Y	Y	NA
Follow dataset or collection	Y	Y	Y	Y	Y
Create new Group	NA			Y	Y
Add dataset to a group (member only)	NA		Y	Y	Y
View/download public datasets	Y	Y	Y	Y	Y
*View/download private datasets			Y*	Y*	Y
*Upload/edit datasets			Y*	Y*	Y
*Delete dataset/resource			Y*	Y*	Y
*Make individual dataset public/private			Y*	Y*	Y
*Make all datasets in collection public/private				Y*	Y
*Change role of collection members				Y*	Y
*Add/delete members from collection				Y*	Y

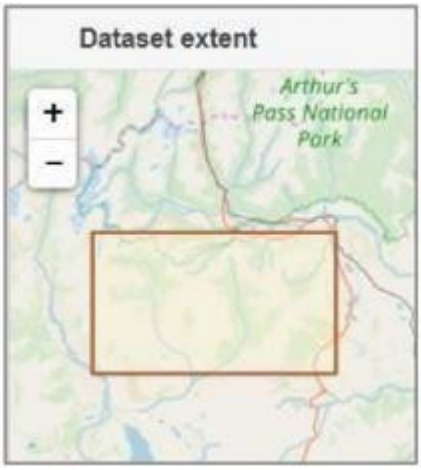
*Edit Collection (Name, Description…)				Y*	Y
Create new Collection				Y	Y
Delete Collection					Y

* = For collections they belong to (or own)

Appendix 2 - The ‘spatial’ custom field

By entering a special custom field name ‘spatial’ you can define the geographic location/ extent of your dataset. This will display a small map to users illustrating the datasets extent and will enable other users to discover your data using spatial searching and filtering.

To enable the spatial feature, you need to provide the spatial extent to your dataset. This is entered in a custom field called spatial - note all lower-case field names - the format of the associated spatial information is geojson.



spatial	{ "type": "Polygon","coordinates": [[[171.3938765526,-43.1762299819], [171.3938765526,-43.0241013542], [171.755897522,-43.0241013542], [171.755897522,-43.1762299819], [171.3938765526,-43.1762299819]]] }

To help, we have some tools to help with this (or the DataStore Administrator can assist):

- We have developed an addin (‘DataUp’) for Excel data which streamlines the process of depositing data to DataStore, including formatting simple spatial extent information.

- We have also setup a simple Excel spreadsheet where the co-ordinates can be entered. Excel will format the required geojson for you to copy and paste into the spatial field in DataStore. Contact the DataStore Administrator for help on this.

To find the co-ordinates, you can either work them out from any geographic fields recorded in your data (e.g., gps points), or you can use the websites below (follow instructions in the Excel spreadsheet). https://boundingbox.klokantech.com (for bounding box outline) https://itouchmap.com/?r=latlong (for multipoint outline)

In future we hope to incorporate the ability to draw an outline of the extent of your data directly on a map within DataStore.

Some examples of some common map spatial extents are given below:

Examples	Type	Formatted text for spatial field
New Zealand (including Stewart Island)	Box	{ "type": "Polygon","coordinates": [[[166.2, -34.2],[179.3, 34.2],[179.3, -47.4],[166.2, -47.4],[166.2, -34.2]]],"properties": {"Geographic coverage": "New Zealand"} }
South Island (excluding Stewart island)	Box	{ "type": "Polygon","coordinates": [[[166.322, -40.3465],[174.4958, -40.3465],[174.4958, -46.6871],[166.322, -46.6871],[166.322, -40.3465]]] }
North Island	Multi	{ "type": "Polygon","coordinates": [[[171.918, -34.794],[173.281, -33.978],[176.972, -36.79],[179.086, 37.659],[178.135, -39.833],[175.815, -41.89],[174.597, -41.639],[174.439, -40.849],[173.179, -39.46],[173.784, -37.535],[171.918, -34.794]]] }
Stewart Island inc Ruapuke	Multi	{ "type": "Polygon","coordinates": [[[167.253, -47.267],[167.391, -47.109],[167.573, -47.019],[167.642, -46.922],[167.561, -46.752],[167.746, -46.64],[168.007, -46.691],[168.572, -46.711],[168.646, -46.791],[168.351, -46.912],[168.26, -47.122],[167.971, -47.176],[167.749, 47.241],[167.551, -47.315],[167.403, -47.309],[167.253, -

		47.267]]] }
Chatham Island inc Pitt	Multi	{ "type": "Polygon","coordinates": [[[-177.175, -44.012],[176.724, -43.547],[-175.885, -43.767],[-176.138, 44.518],[-177.175, -44.012]]] }