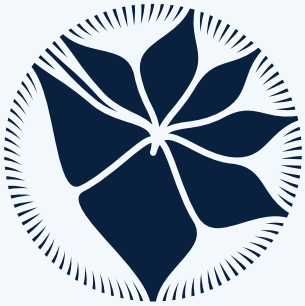


ChecklistBank

ChecklistBank Project

A tool to build and curate checklists

Catalogue of Life

Version 925fd12, 2025-01-16 11:51:31 UTC

GBIF Global Biodiversity Information Facility

Table of Contents

Description

ChecklistBank is a platform that supports the publication, analysis, and curation of checklists with emphasis in taxonomic and nomenclatural data. It is an integral part of the infrastructure developed and maintained by the Catalogue of Life (COL) and the Global Biodiversity Information Facility (GBIF).

A detailed tutorial of the general tools of ChecklistBank can be found here.

In addition to the general tools, users can create and publish new datasets through ChecklistBank in two different ways:

- Project: datasets constructed and managed within ChecklistBank
- External: datasets created outside ChecklistBank and imported as FAIR data resources into the platform.

Both of them benefit from many common features that ChecklistBank delivers as an open data infrastructure: standards compliance, validation and reporting, multiple download options, rich API support, etc.

ChecklistBank users can use the **Project functionality** to create new Checklists from existing datasets hosted in the platform (also referred here as sources), combining their data totally or partially. For example to create a national checklist, based on taxonomic checklists and articles from a specific country.

This tutorial describes the functions available for assembling a checklist and delivering it as a dataset. These tools are already in use to construct the Catalogue of life data product.

Project functionality includes the following features:

- Interactive assembly process.
- Include total or partial information from any other datasets in ChecklistBank.
- Select or block specific parts from a any given dataset included in the project.
- Synchronisation between the main project and the source dataset.
- Stable static versions called releases, with stable record identifiers.

Audience

This tutorial is for Biodiversity information systems, taxonomists, governmental entities, online platforms, and any other people or institutions interested in assembling a checklist from preexisting lists in ChecklistBank.

ChecklistBank login

All registered users on GBIF can login on ChecklistBank:

1. Go to https://www.checklistbank.org/
2. Click on Login (top right)

3. Login with your GBIF account

If you don't have a GBIF account yet, use the register at gbif.org now!

Use case

In this tutorial you are going to explore some of the features available in ChecklistBank by creating a new Moth dataset. You will use four Lepidoptera (moth and butterfly) datasets available in ChecklistBank. The three-letter abbreviations GLI, ALU, PHR and GAR will be used through the rest of this tutorial to refer to the following datasets:

- GLI: Global Lepidoptera Index – the primary dataset used in the Catalogue of Life Checklist for most groups of moths and butterflies
- ALU: Catalogue of the Alucitoidea of the World – the dataset used in the Catalogue of Life Checklist for the many-plume moths (superfamily Alucitoidea)
- PHR (id: PLZ31420) A world catalogue of Phragmataecia (Lepidoptera: Cossidae), with a new species from Kazakhstan and Kyrgyzstan – a dataset summarising the names currently accepted for a moth genus, extracted by Plazi from a paper in the journal Zootaxa
- GAR (id: PLZ41041): A review of the genus Gargela Walker in China, with descriptions of ten new species (Lepidoptera: Crambidae, Crambinae) – a dataset summarising the names of Chinese species a moth genus, extracted by Plazi from a paper in the journal Zootaxa

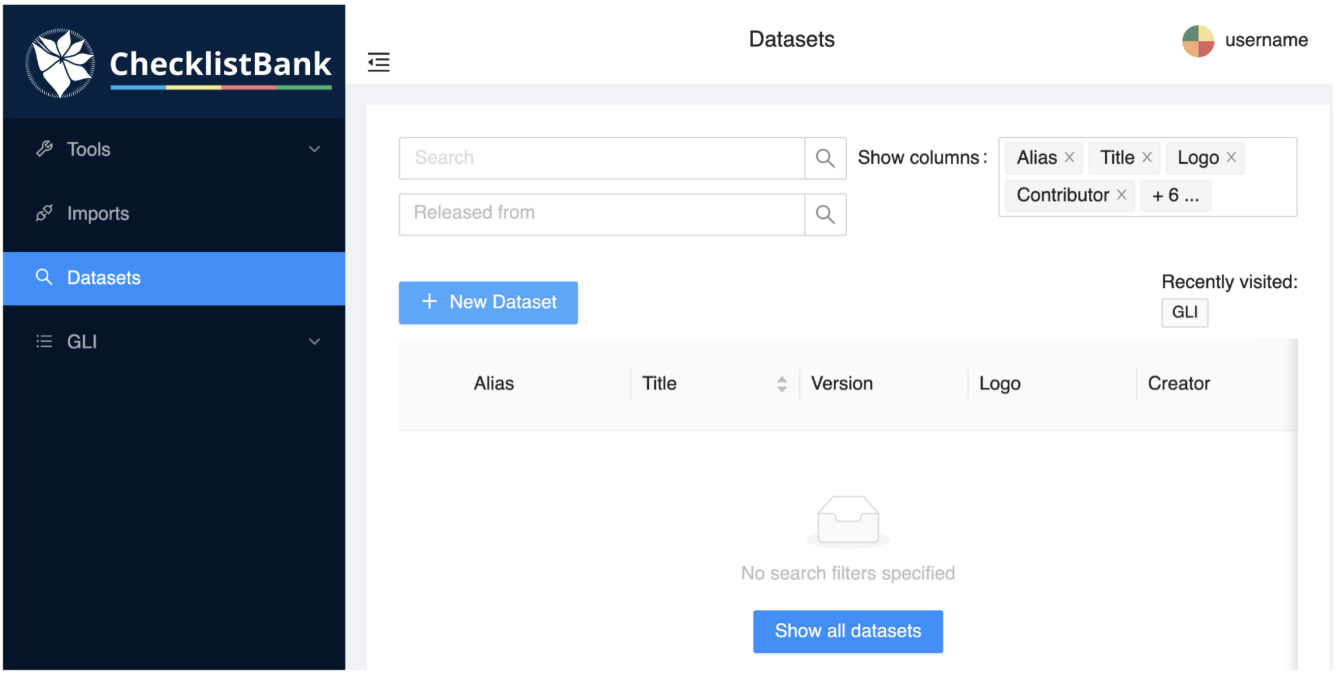
You will use ChecklistBank tools to structure a new moth dataset based on GLI and then modify it in the following ways:

- Exclude the butterfly superfamily Papilionoidea, to create a "moth"-only dataset.
- Replace the Alucitoidea section in GLI with the version in ALU .
- Replace the Phragmataecia section in GLI (in the family Cossidae) with the version in PHR.
- Merge new species from GAR with those already in the Gargela section in GLI.

The tutorial will finally show how the resulting dataset can be exported.

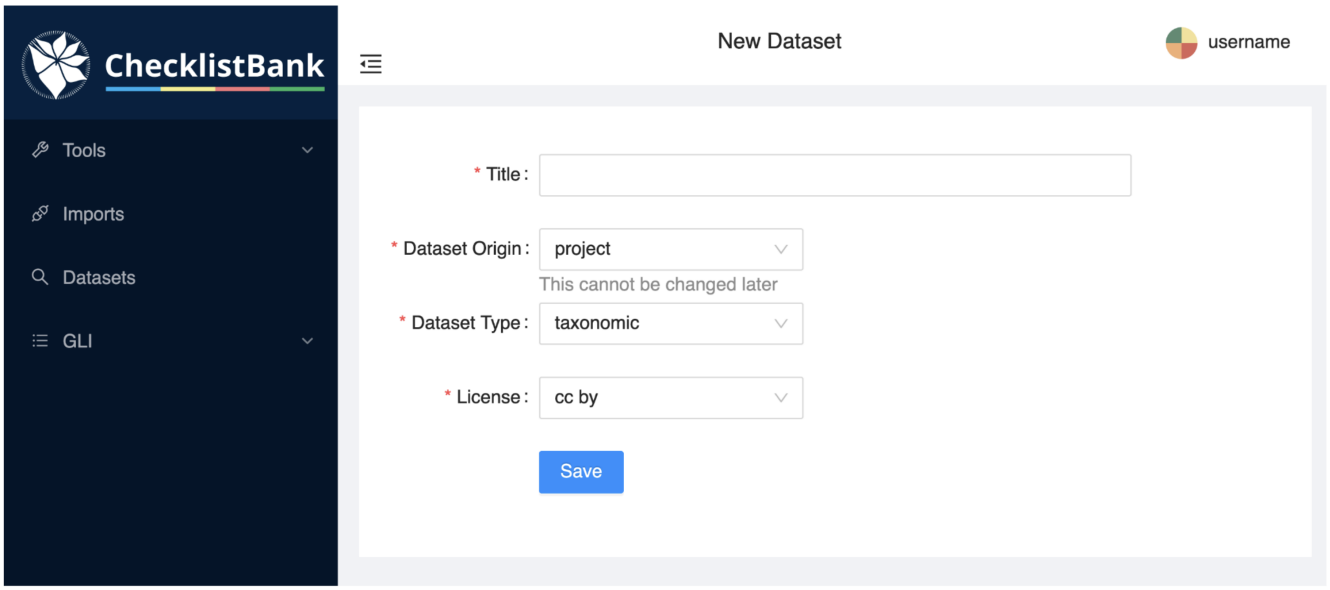
1. Creating a new project

Go to Datasets on the menu on the left side of the ChecklistBank window, and then click on '+New Dataset'.



Complete the form:

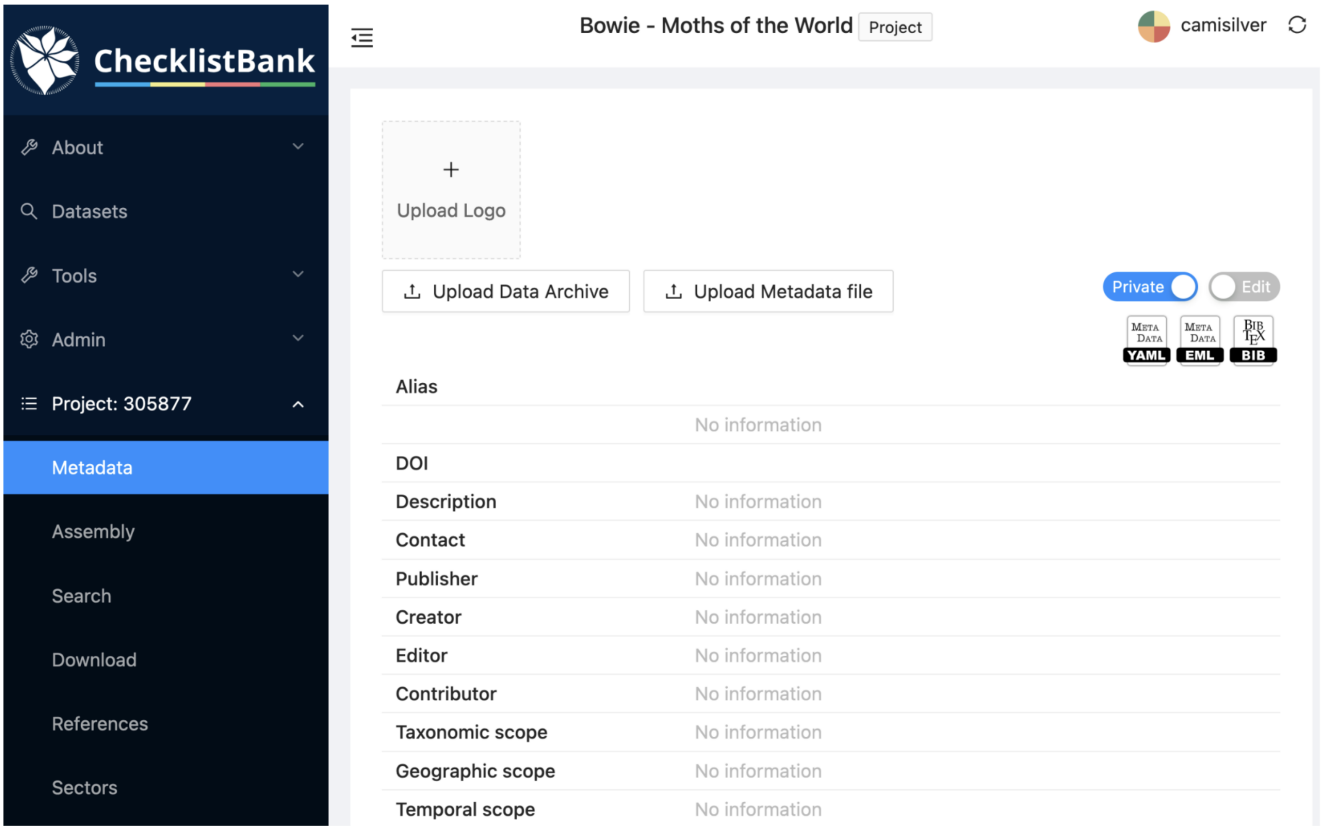
- Set a meaningful and clear title. For this tutorial use: your name followed by "Moths of the World".
- Select "project" for the Dataset Origin and "taxonomic" for the Dataset Type.
- Set a Creative Commons licence for the project: “cc by” is appropriate since none of the four source datasets of this tutorial have more restrictive licences. It is important to consider that the licences of the source datasets must be compatible, that is, if the base source has a less restrictive licence, it will not be possible to integrate information from a more restrictive source. The licence of the project can be changed at any time.
- Click 'Save'.



2. Metadata

Once created the project, ChecklistBank will display the overview metadata page. Click on 'Edit' at the top right to add metadata. Fill in as much detail as possible, specially in the following fields and taking into account the following best practices:

- Give the project your own unique Alias – this will make it easier to locate in later steps. Consider using camelCase and spaces; avoid ALL CAPS and underscores.
- Description: A brief summary of the content and scope of the dataset.
- Contact: Person who can answer questions regarding the content of the dataset.
- Creator: Creator of the dataset.
- Contributors: All those who contributed to the creation of the checklist. Give as many details as you want.
- Version: to be updated every time updates are made.



Then click 'Save' at the bottom of the page.

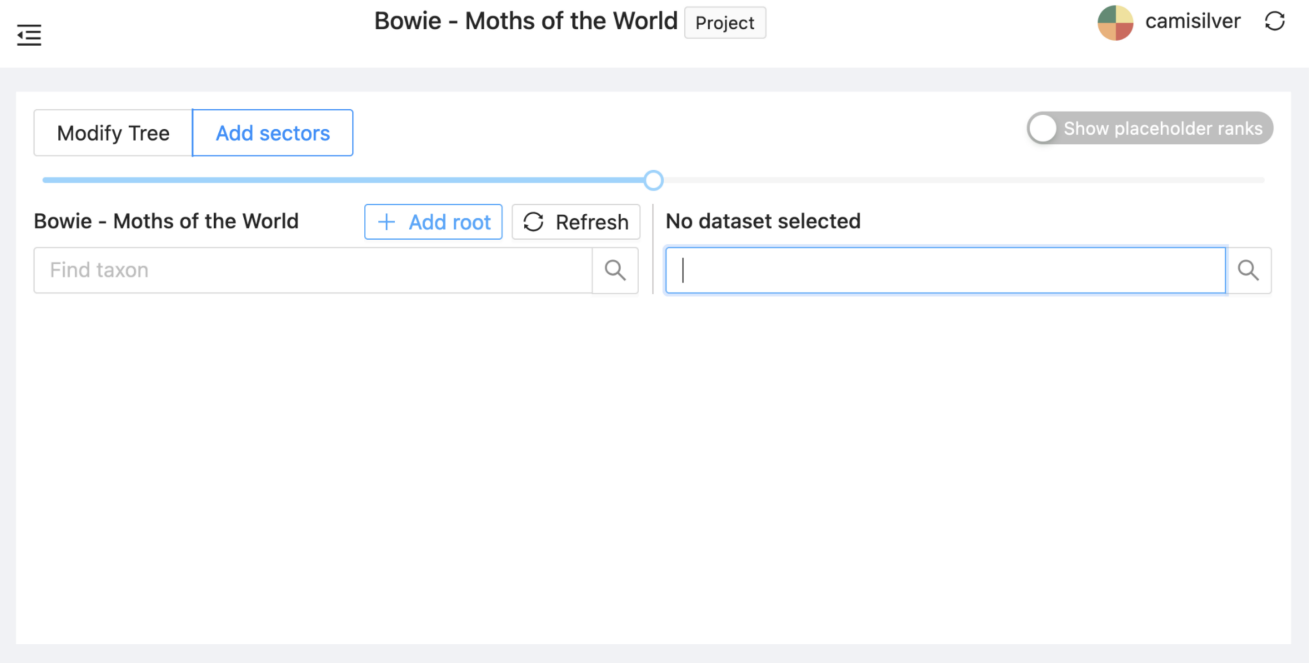
Once the Project is created, its Alias will be visible in the menu bar at the left, under which are all the sub items of the project.

3. Checklist assembly

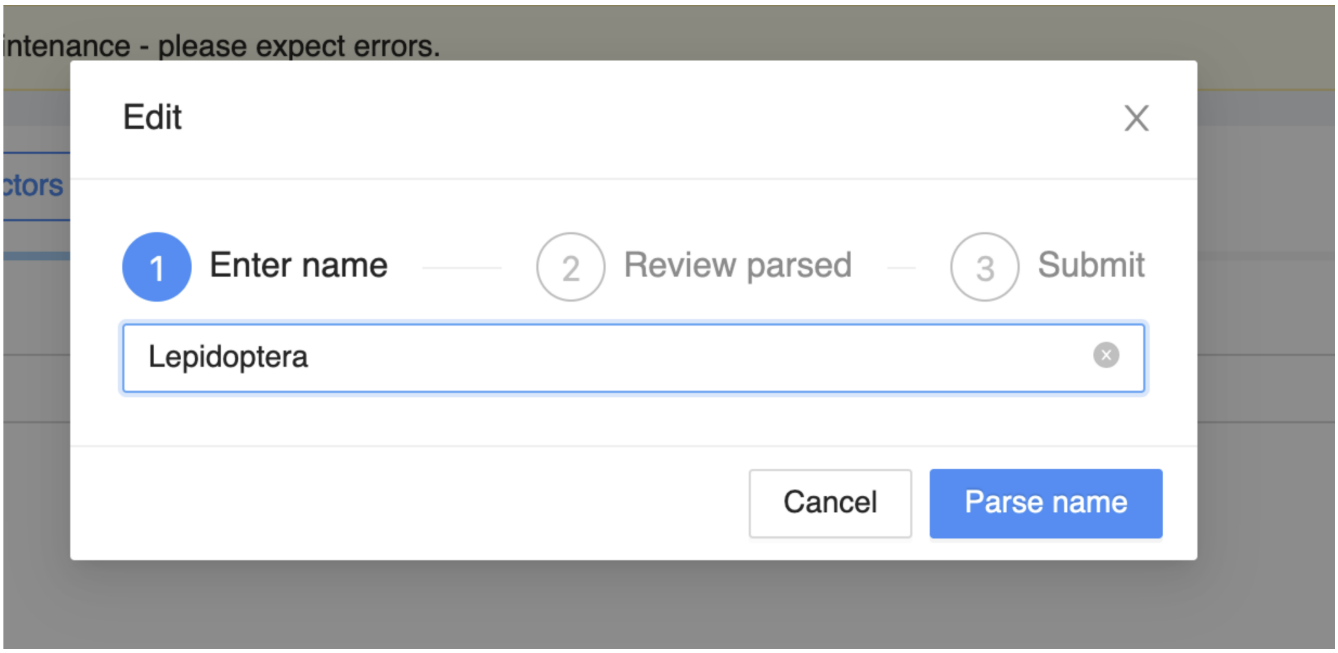
Building a new checklist consists of several steps and tools, which are described below:

3.1. Add a root taxon

To start the checklist assembly the highest taxonomic level has to be established. Under the Alias of the project on the left menu select the Assembly subitem, click on the '+ Add root' button near the middle of the window.

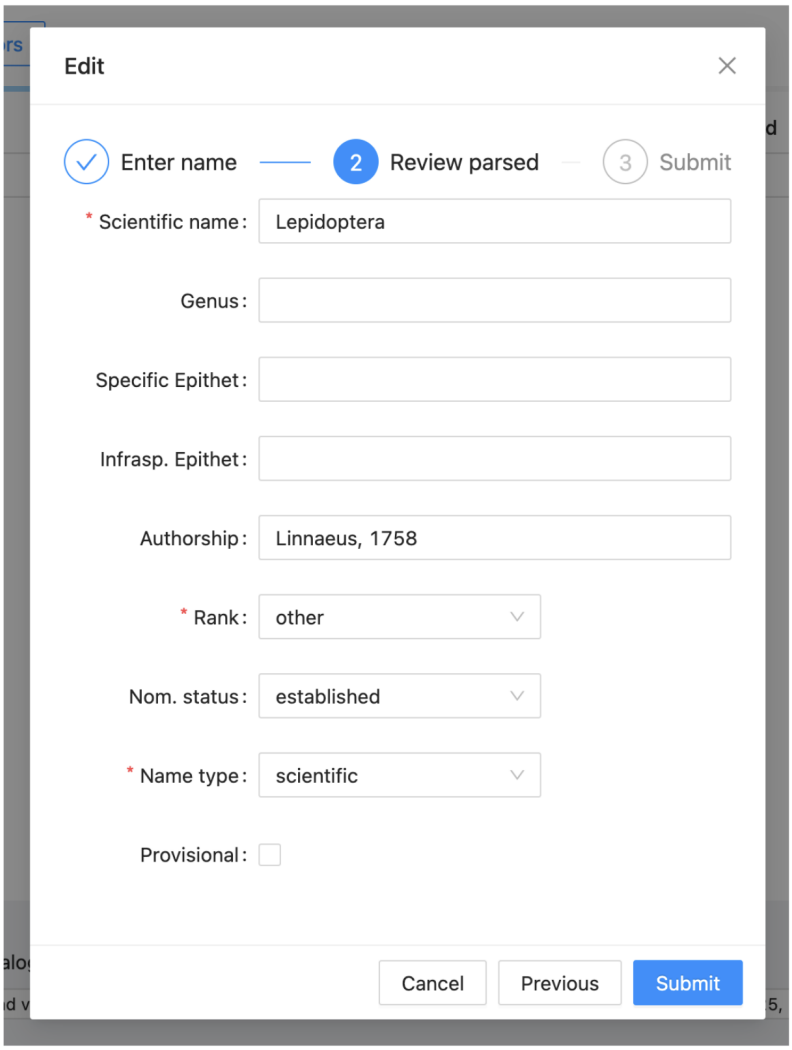


Enter the name “Lepidoptera” and click on Parse name.

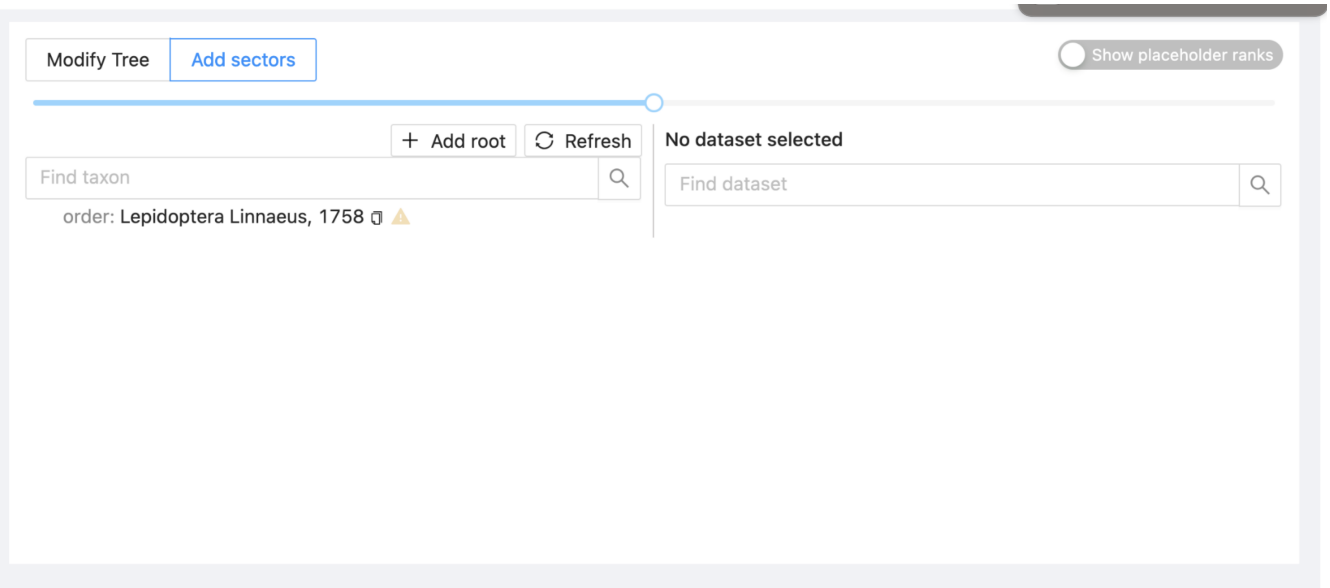


A form will appear, change the Rank to order. This will change the display to include the fields shown in the image. Complete these as follows:

- Name status: established
- Authorship: Linnaeus, 1758



Then click on 'Submit'. The project now includes a single taxon (the order Lepidoptera) with a yellow triangle indicating that at present no sectors have been added.



3.2. Add and manage sectors

The building blocks for a ChecklistBank project are called Sectors, these can be created by adding an entire dataset or a specific taxonomic group within a dataset. You can create as many sectors as

needed, this gives maximum flexibility to assemble the project. There are several methods to add sectors to the assembly:

- **Union**: adds all the descendants of the subject taxon, and ignores the name of the subject itself
- **Attach**: adds the subject taxon and all its descendants
- **Merge**: enriches the existing assembly with additional information provided by the selected sector like: new names, synonyms, authorships, and references, among others.

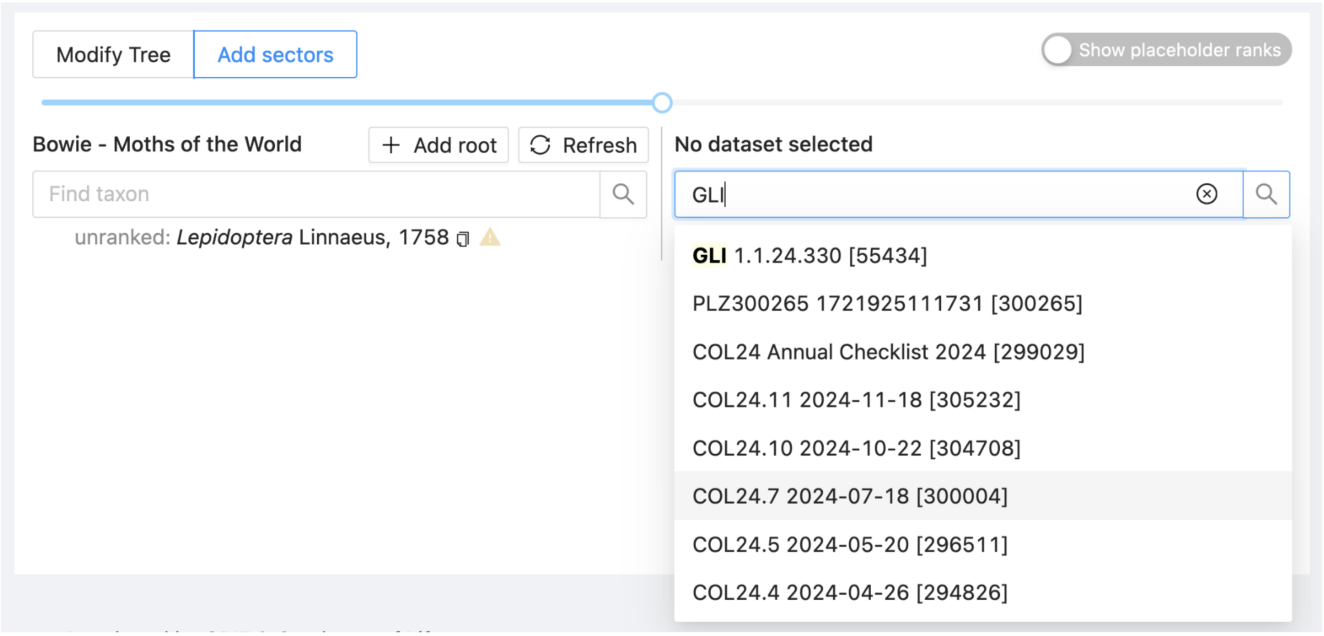
There are also several editorial decisions that can be made over any taxon on the project:

- **Block**: ignores the selected name and its children.
- **Ignore**: only the selected name is ignored.
- **Review**: indicates that the selected name has already been reviewed and requires no changes

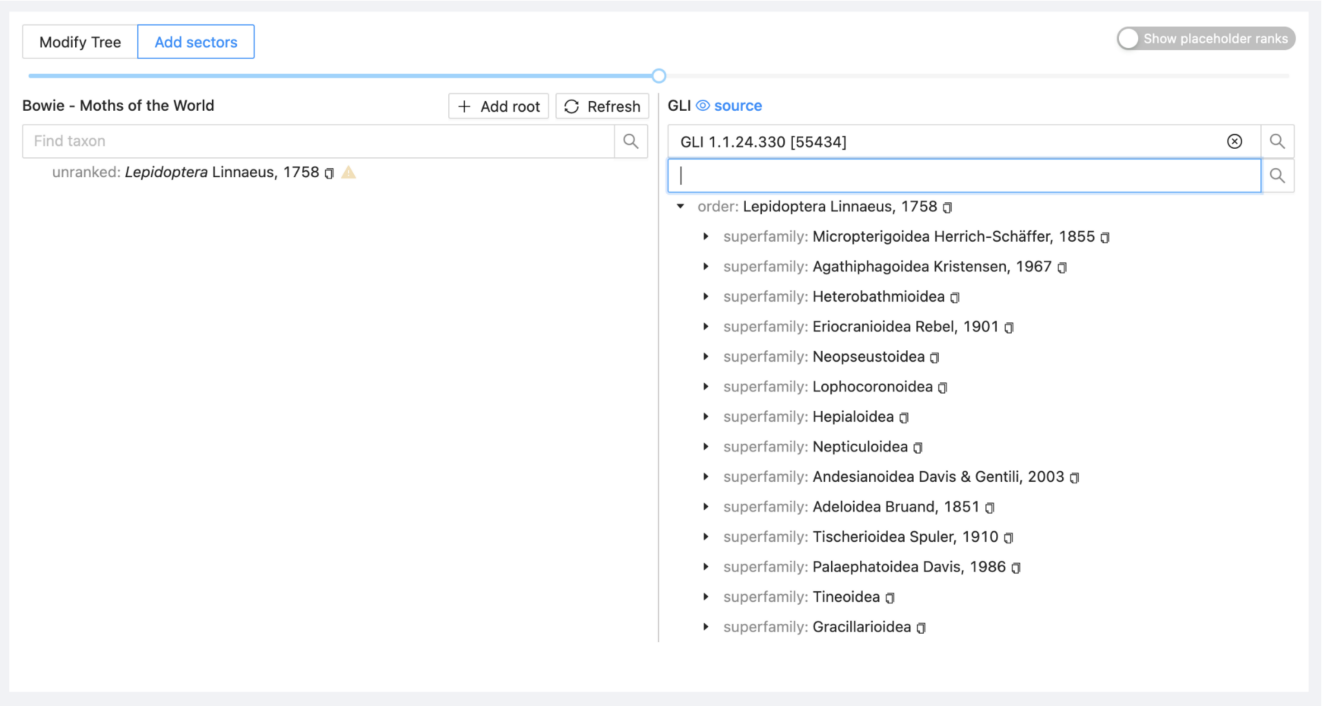
3.2.1. Union

You will start by adding the base tree, GLI, as the first sector with the union method. Make sure that the 'Add sectors' option is selected at the top of the window.

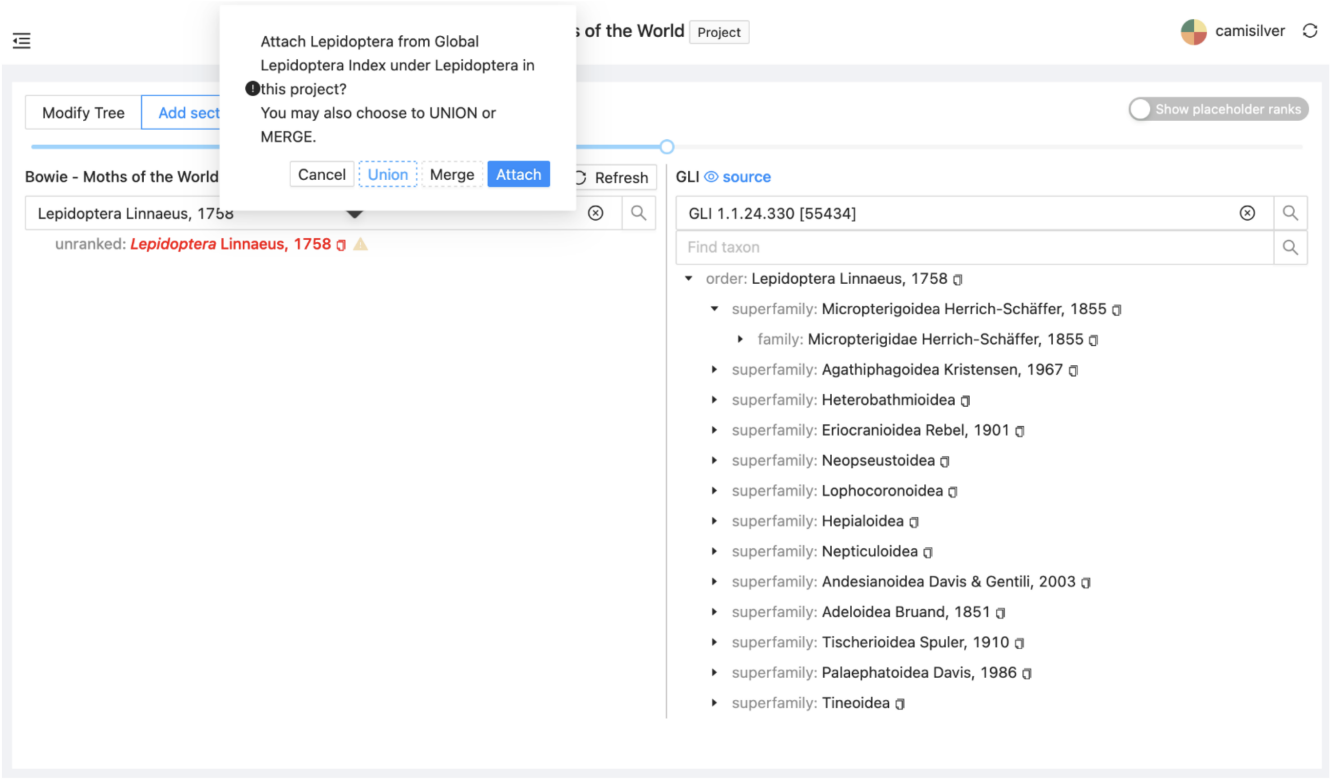
On the search box on the right side of the window, enter “GLI” or “Global Lepidoptera Index”. Alternatively, type “55434”, which is the dataset ID for GLI. Wait a few seconds for the dropdown list of datasets to load.



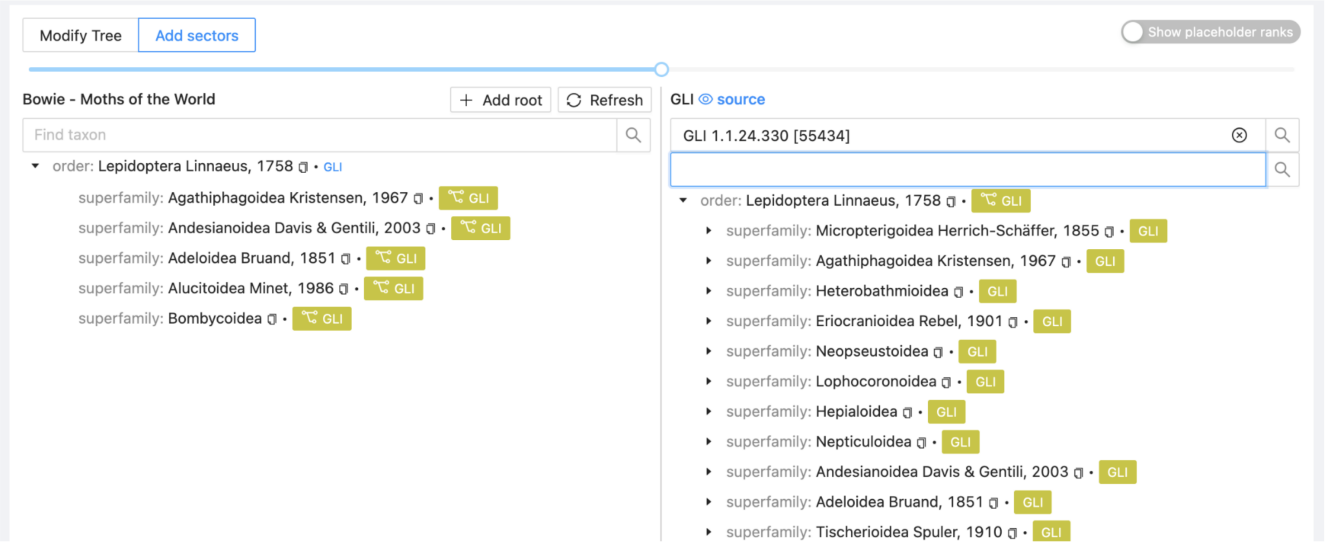
The right side of the window will refresh with the hierarchy for GLI.



Select Lepidoptera on the right and drag it across onto or immediately below Lepidoptera on the left. You will be asked to specify how the GLI Lepidoptera should be added as a project sector. Select 'Union'.



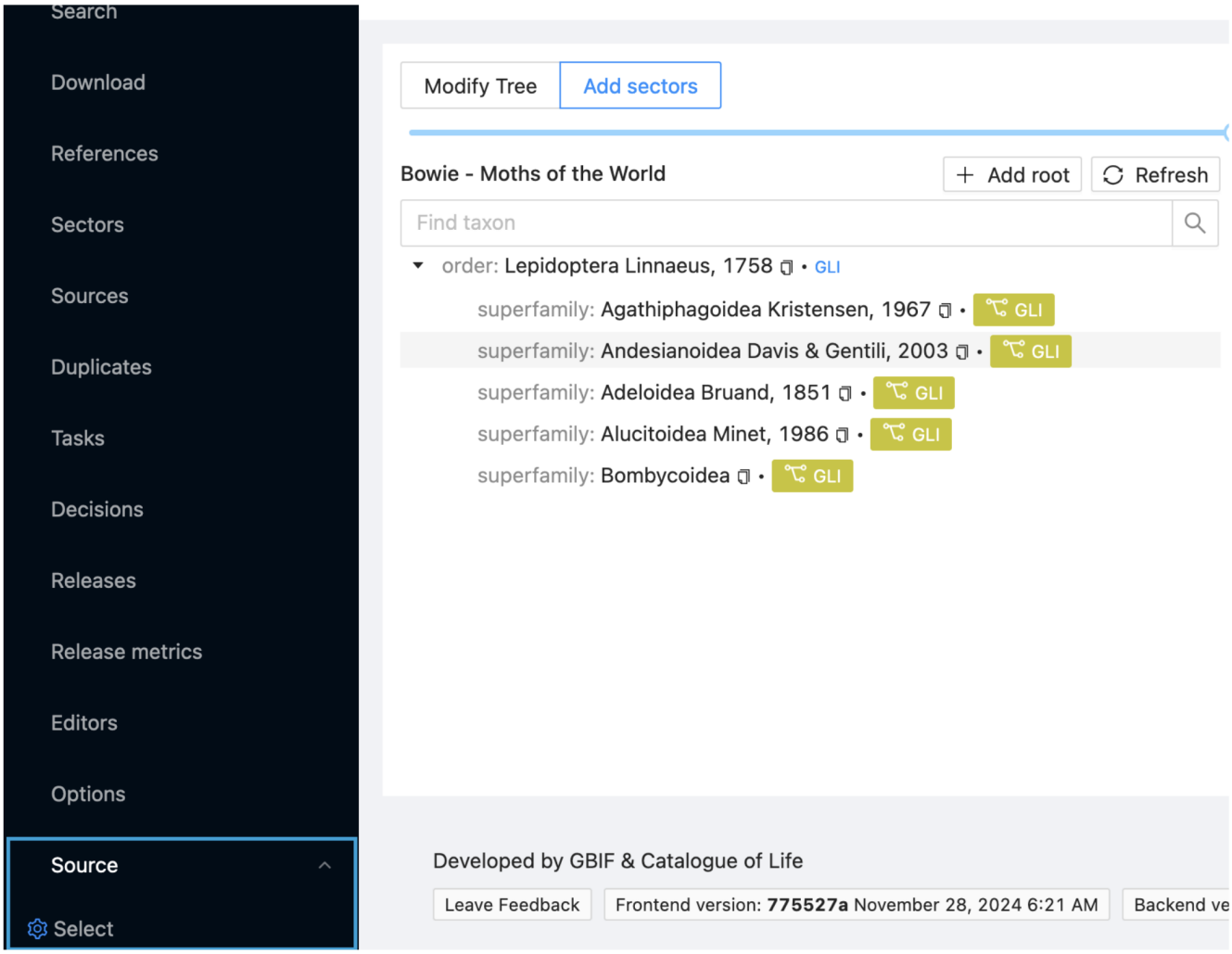
The left side of the window will refresh to show that Lepidoptera is now sourced from GLI and will list just the first five children of GLI as an indication of content. The coloured labels will allow GLI to be distinguished easily from other sectors as they are added.



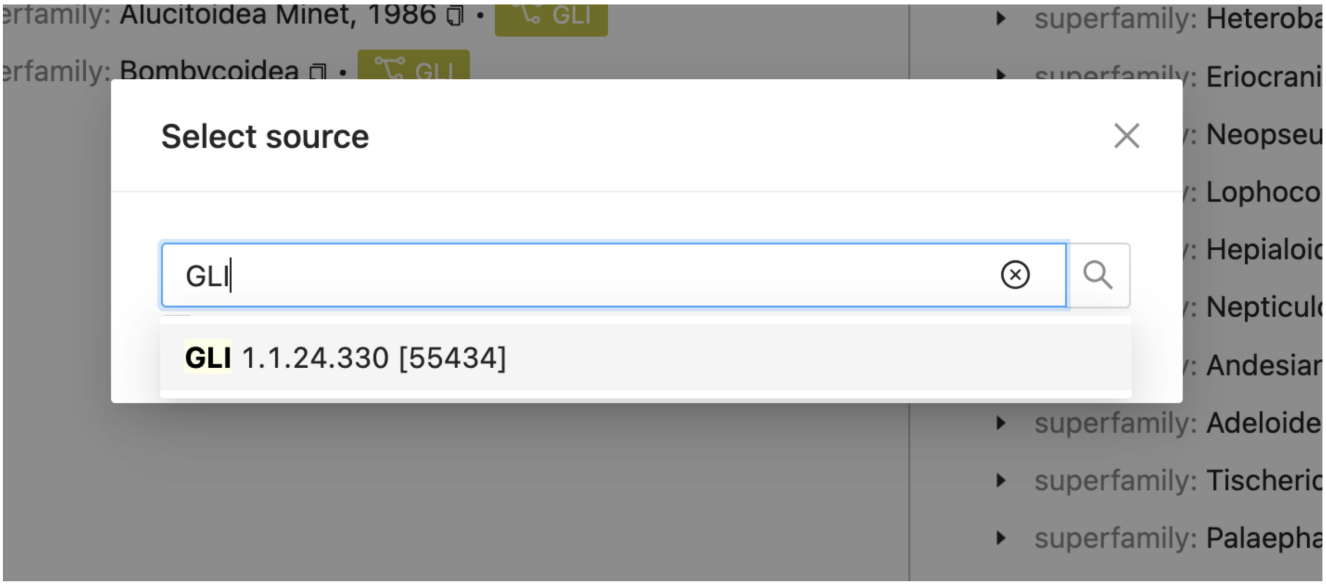
3.2.2. Block

Now you are going to add a decision to exclude some parts from GLI: the superfamily Papilionoidea, the superfamily Alucitoidea and the genus Phragmataecia.

Click on the 'Source' item in the menu at the left of the ChecklistBank window. This will open a Select option with a second blue cog. Click on the cog.

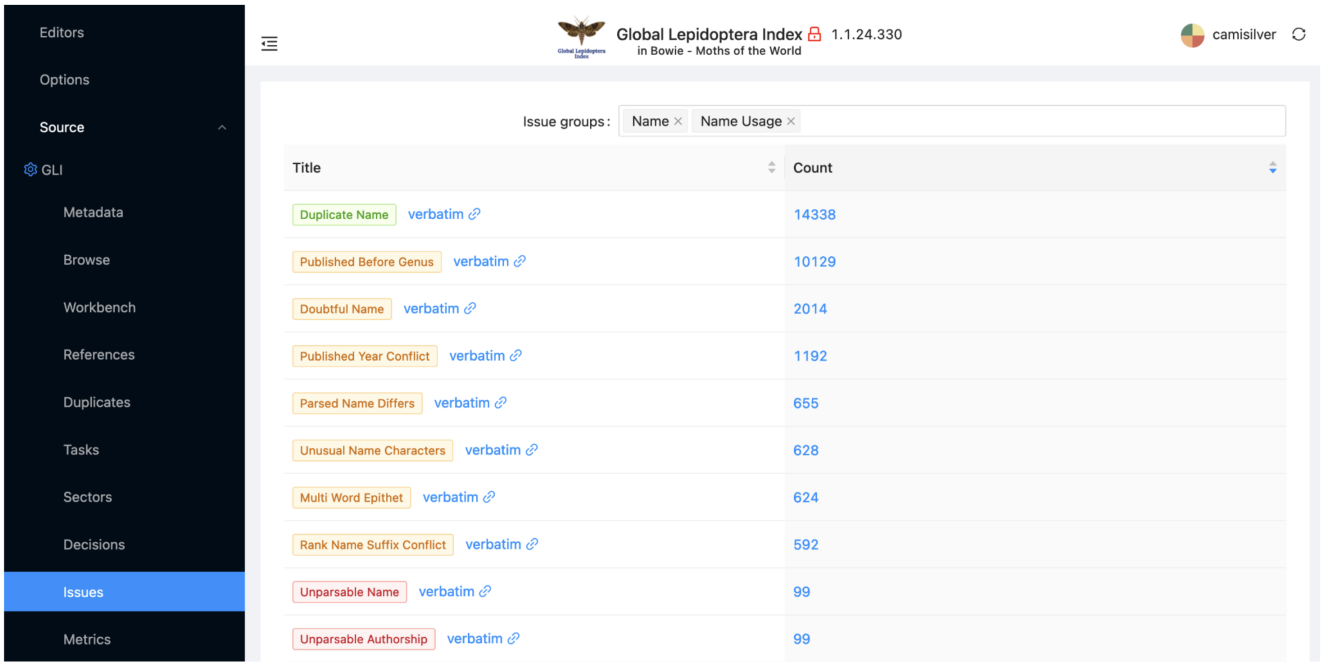


You only have one sector in your project (GLI). Select this from the pop up search window.

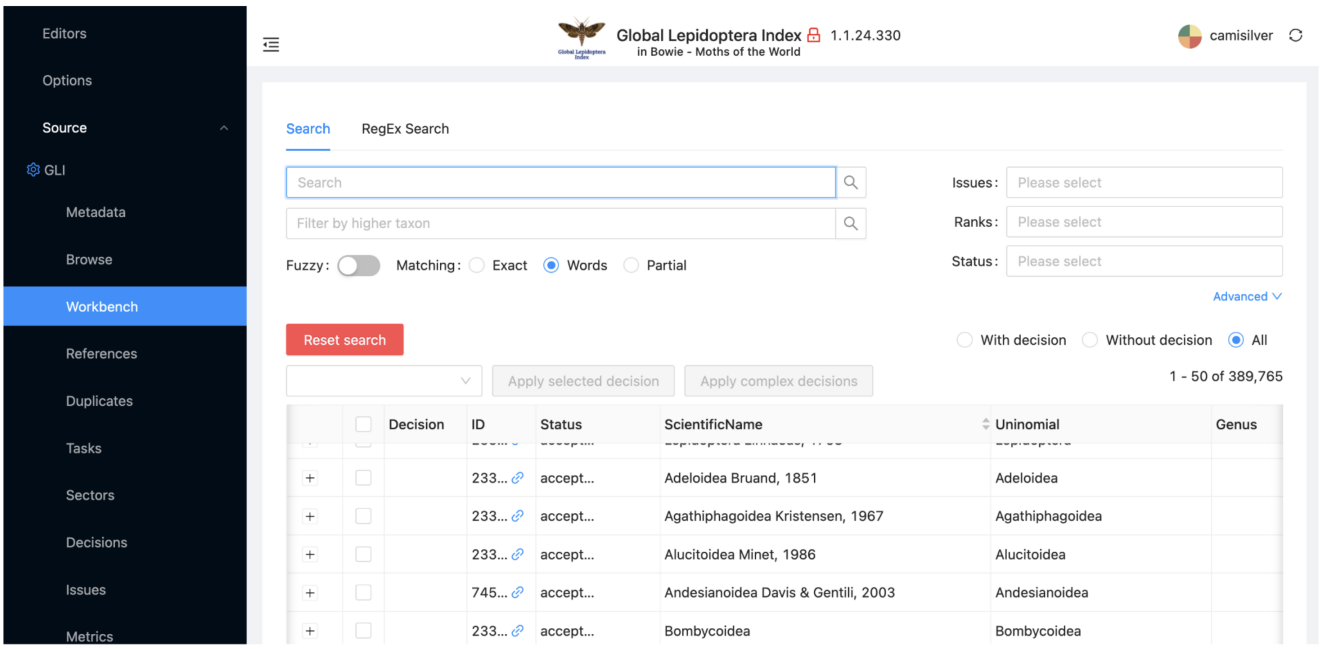


A new view of GLI will appear next to a blue cog, representing the active source dataset for workbench operations. As more sectors are added, clicking the cog again will allow you to switch the source dataset.

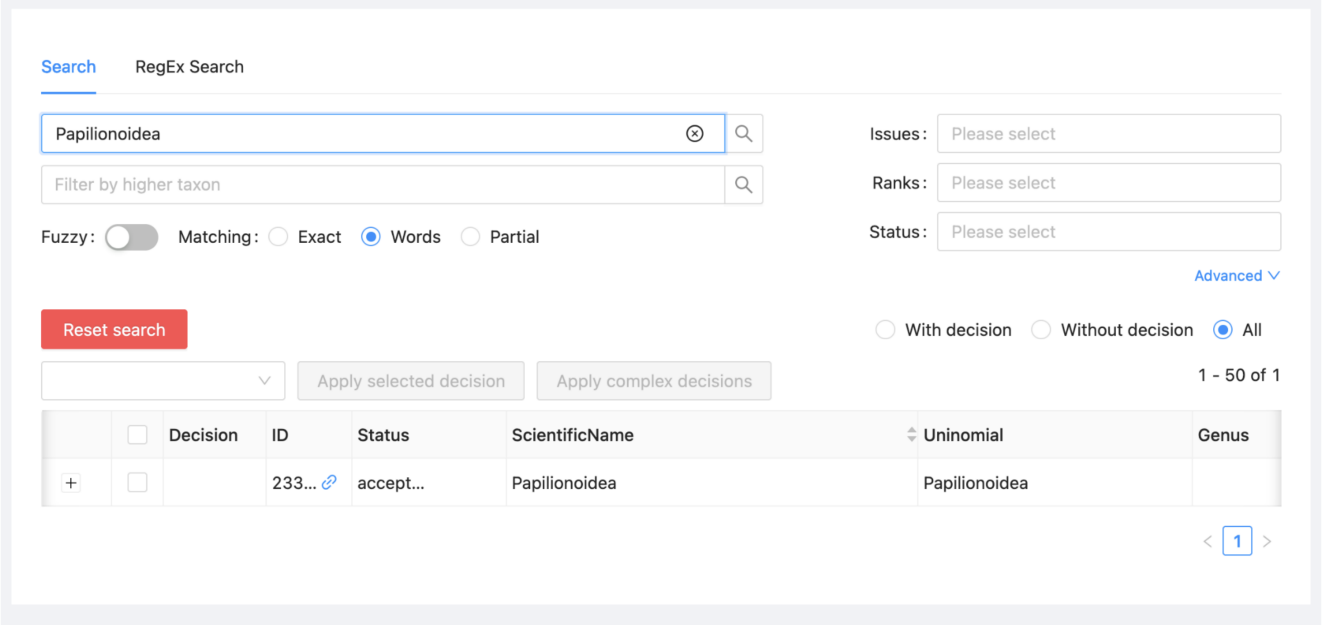
The initial view displays the issues identified for the dataset. Since GLI is very large and actively curated, there are many issues, but they can be ignored for this tutorial.



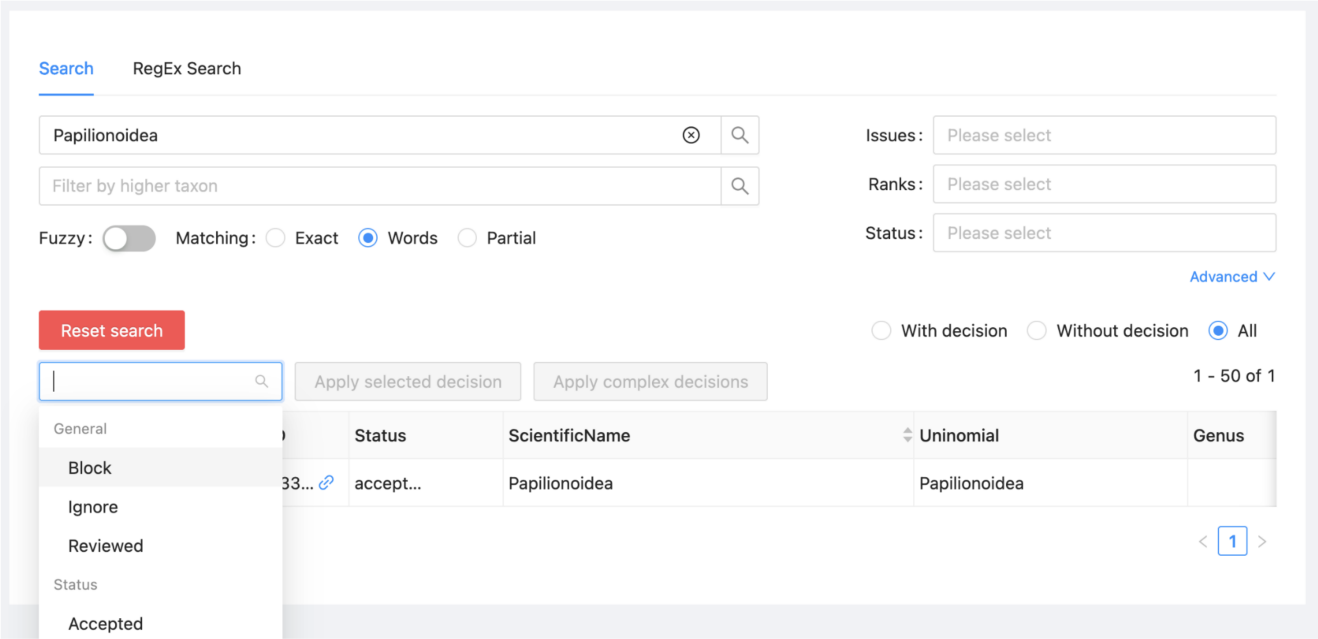
Select the 'Workbench´ subitem for the Source view of GLI. The workbench shows all of the taxa that will be included in your project dataset from GLI.



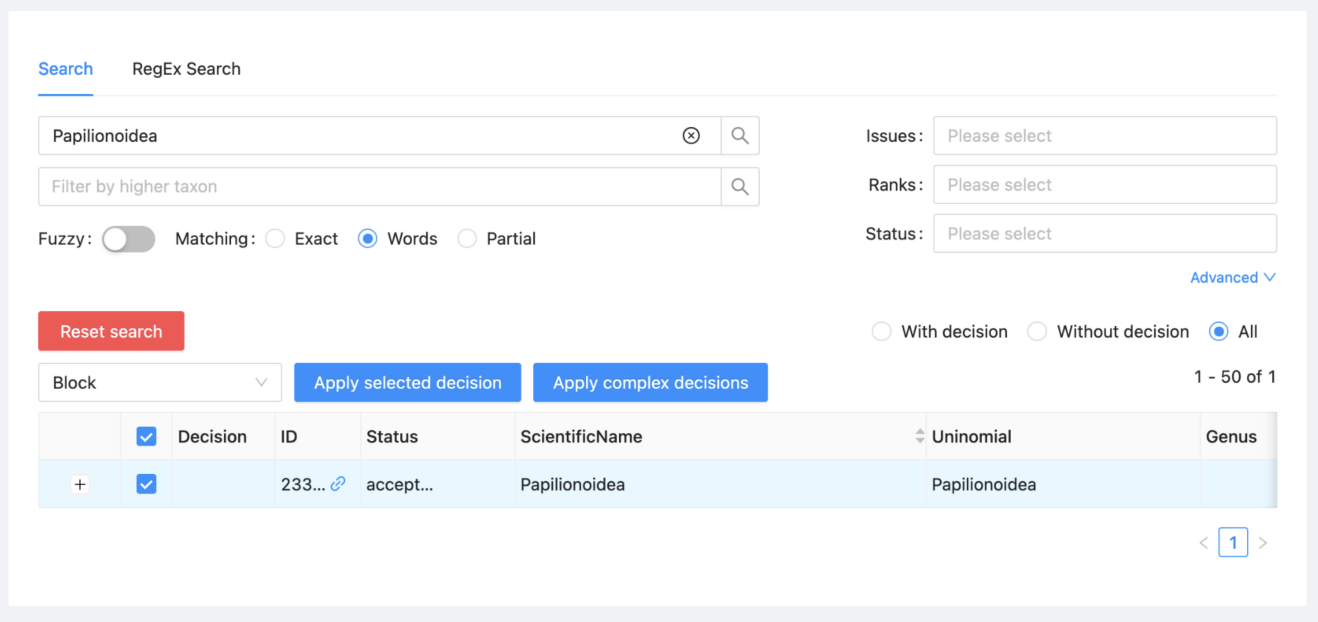
Type “Papilionoidea” in the Search box at the top of the window and hit Enter. The record for the butterfly superfamily Papilionoidea will be the only one shown. You can alternatively filter taxa using ranks or higher taxon or other options that can be selected by clicking on 'Advanced'.



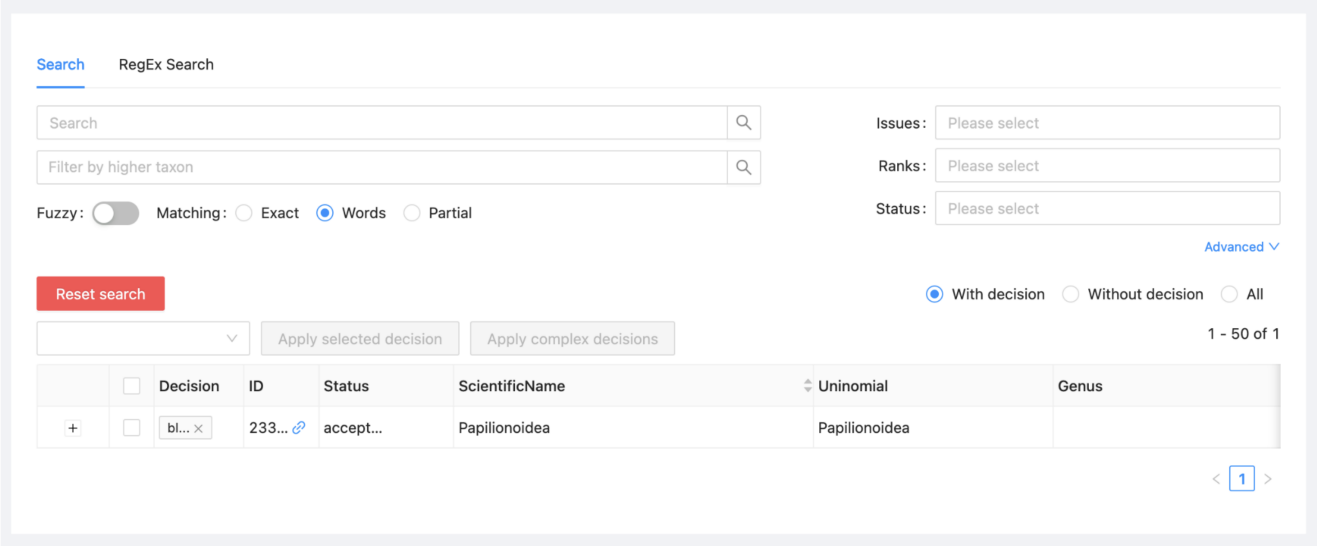
Select "Block" from the dropdown at the left above the taxon list (just below “Reset search”).



Select "Papilionoidea" from the list and click on 'Apply selected decision'.

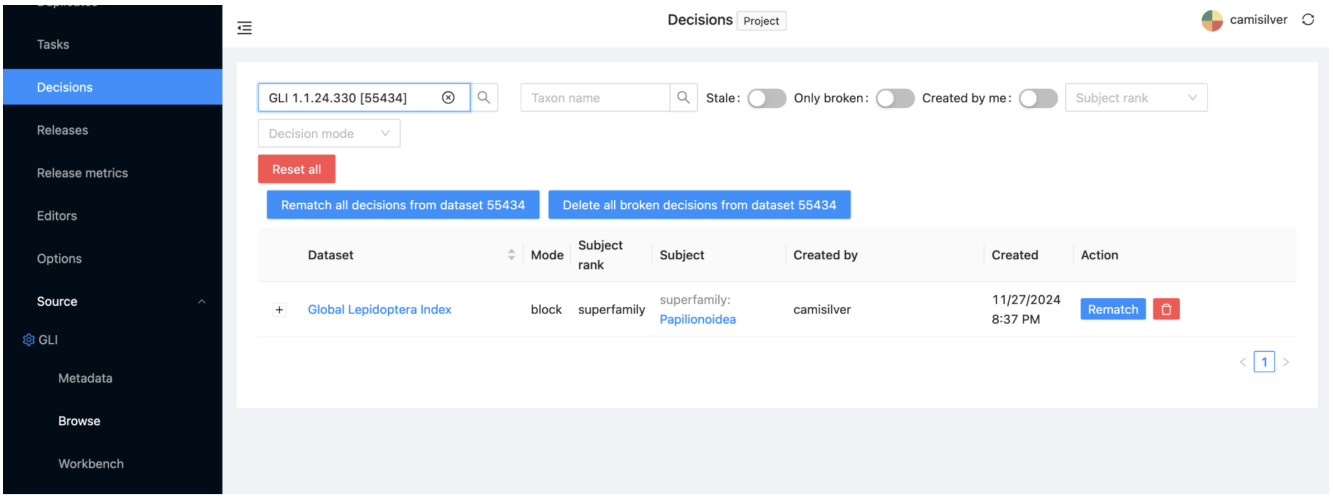


If you click now on ‘Reset search’ (the red button on the left) and the ‘With decision’ radio button on the right, you will see that you have now specified a single decision, that Papilionoidea should be blocked. The same information can also be seen if you click on the Decisions subitem under the project menu.



When you populate the dataset, this decision will be honoured.

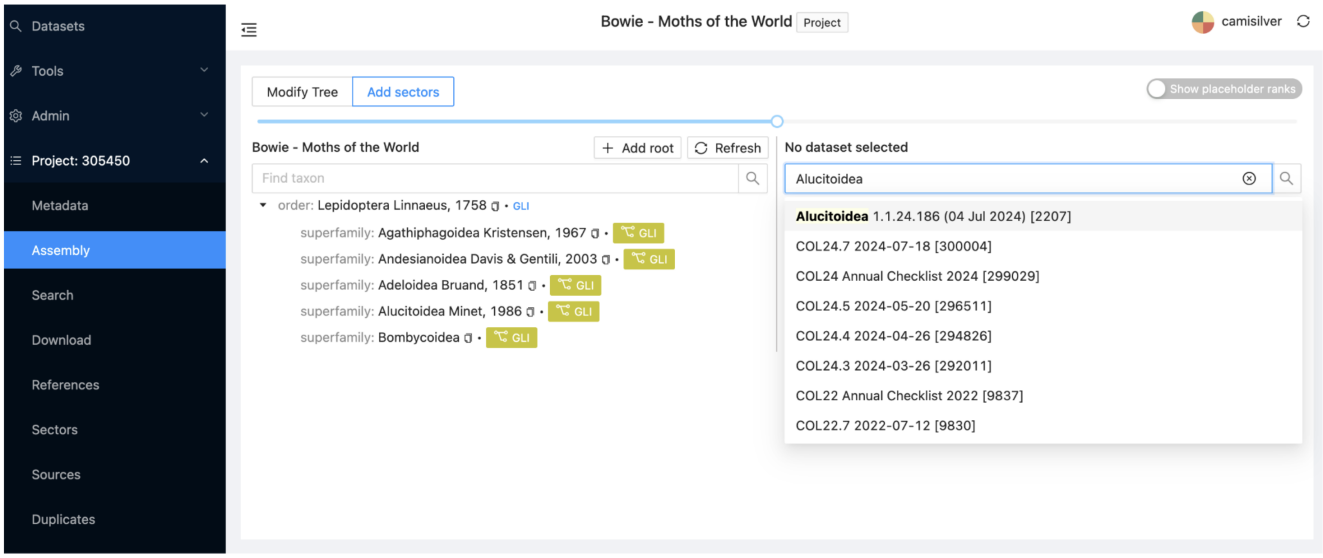
Repeat the same steps (Source: Global Lepidoptera Index > Workbench > search for taxon > Block taxon) for the superfamily Alucitoidea and for the genus Phragmataecia.



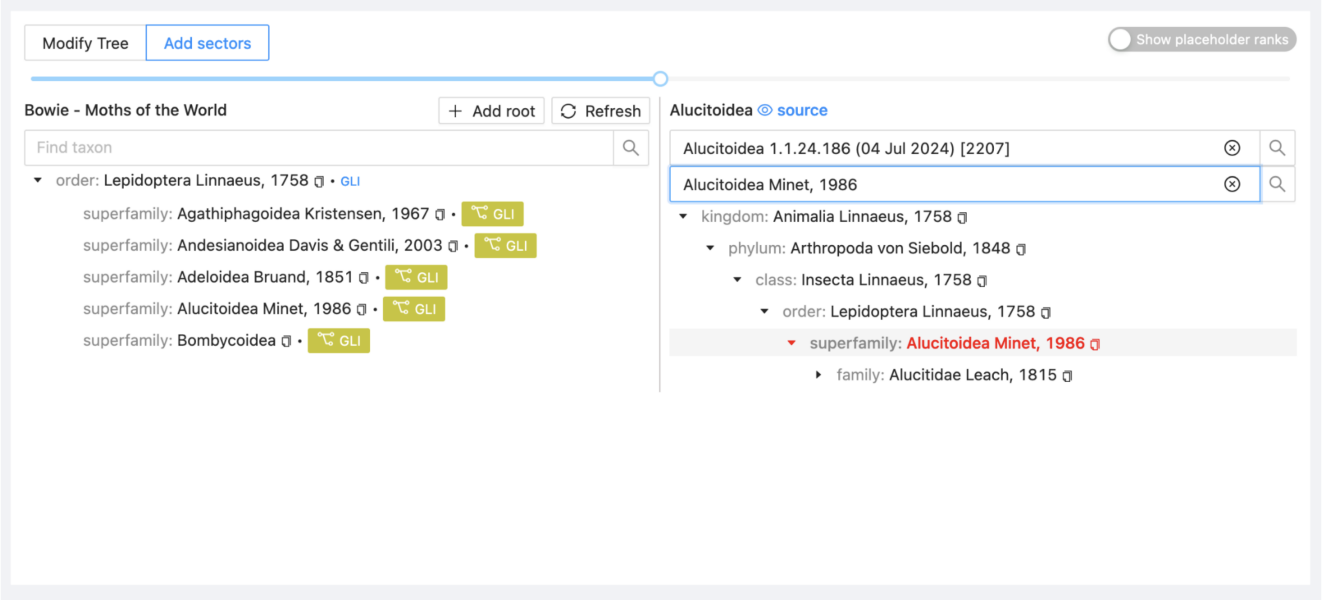
3.2.3. Attach

Unlike GLI, the dataset Catalogue of the Alucitoidea of the World (ALU) includes higher classification all the way up to the kingdom, thus you are going to replace GLI Alucitoidea sector with ALU source.

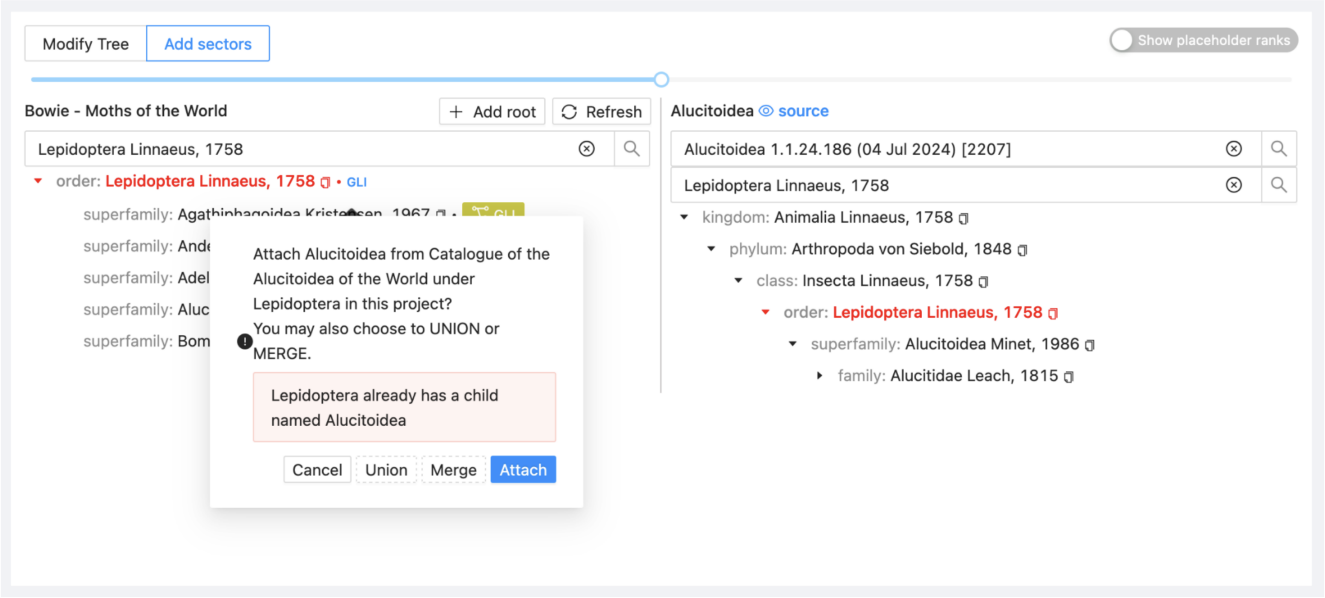
In the project, select the Assembly subitem at the left and make sure the Add sectors view is selected (at the top of the window). Enter “Alucitoidea” in the search field on the right, wait a couple of seconds for the drop down list of datasets to appear, and select the first dataset (id:2207).



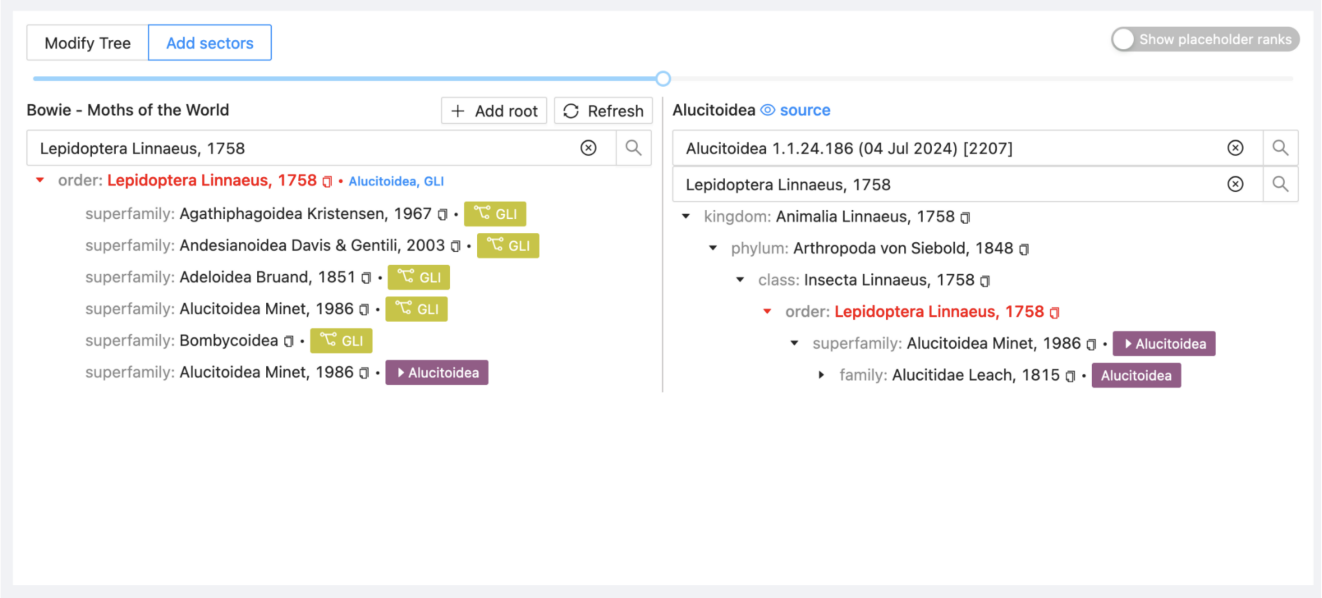
Select the small arrows next to each name until the superfamily Alucitoidea appears. Alternatively, type “Alucitoidea” in the second search box on the right of the window and select it directly.



Drag Alucitoidea from ALU on the right onto Lepidoptera in your project on the left. Select Attach to include it as a child taxon.



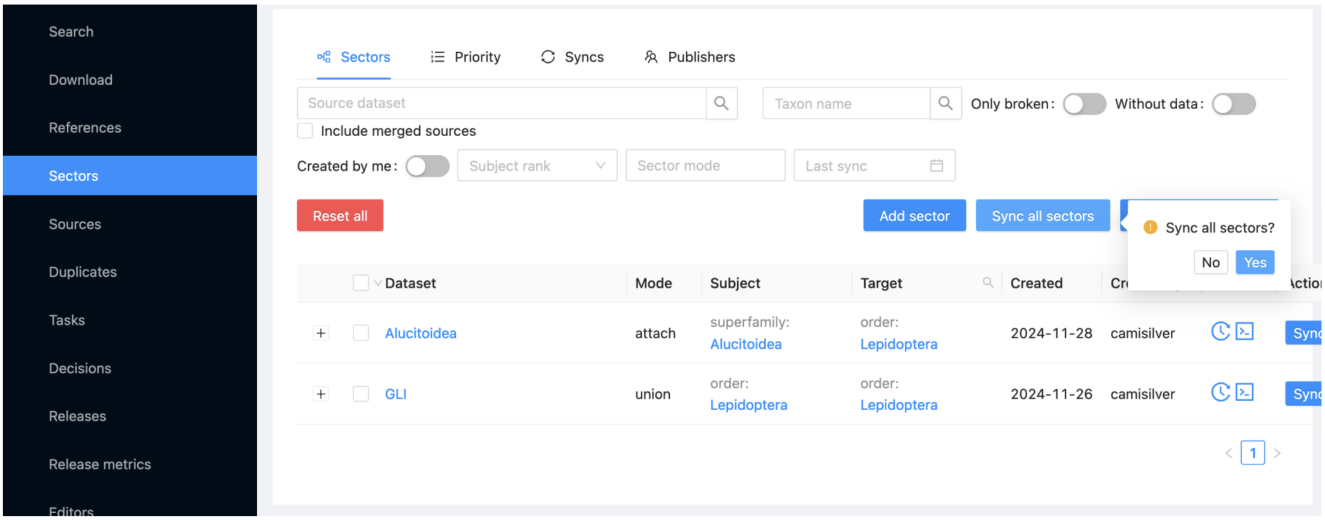
The project display will now show that you have added two sectors. Note also the blue eye symbol and source link at the top right. This is a quick way to select the current dataset as the current source dataset in the Source submenu.



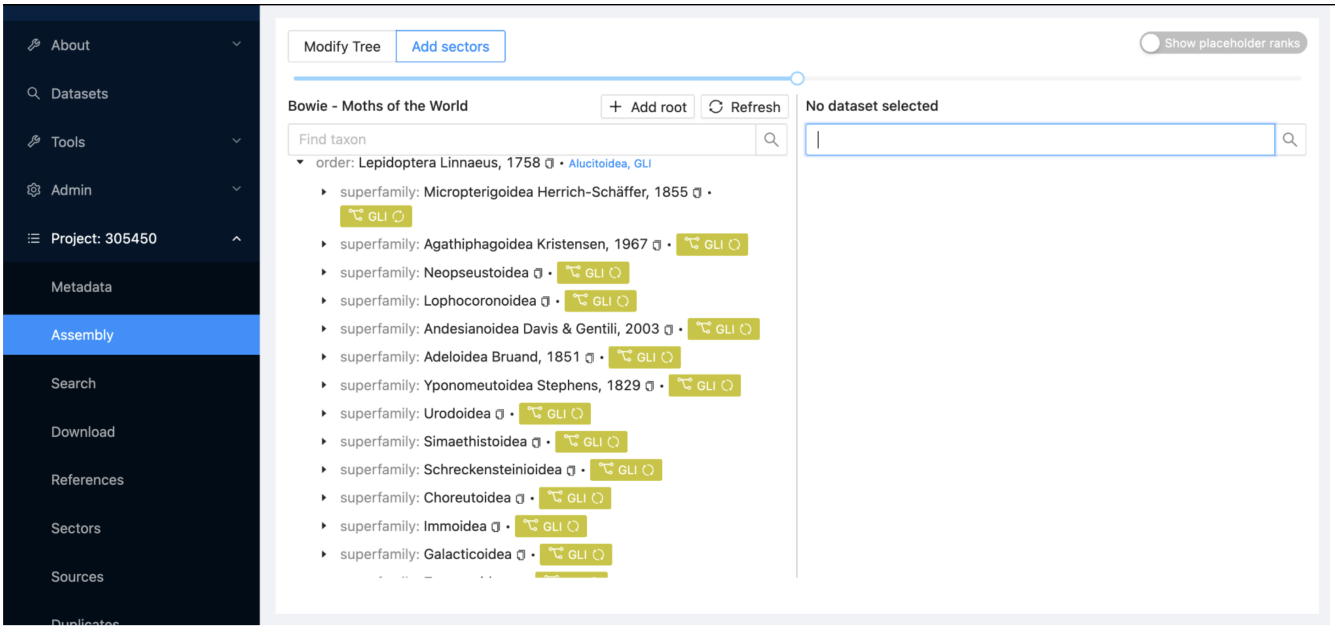
At this point the tool only shows an overview of the higher classification, to continue creating sectors at a more specific level it is best to synchronize all sectors so the entire assembly can be explored and managed.

Select the Sectors item in the menu at the left below the alias of your project. This view allows you to synchronise the content from each sector into your project.

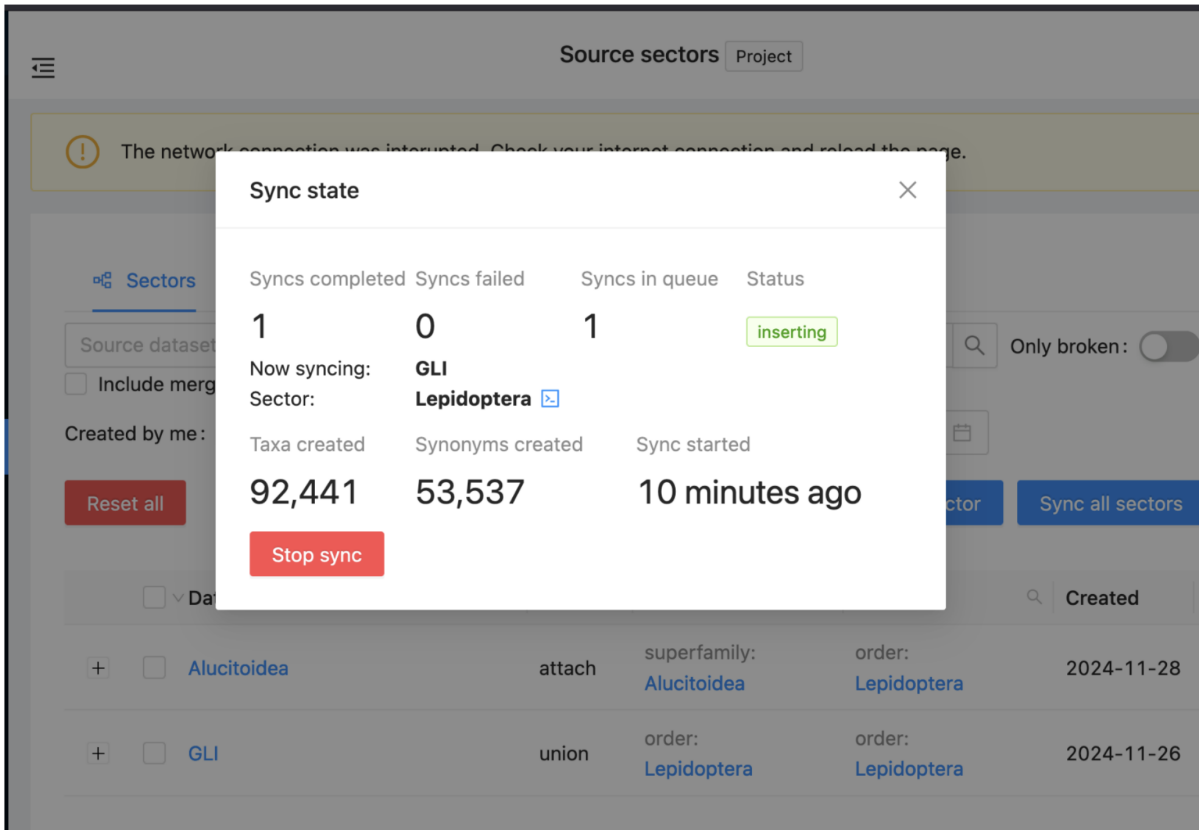
Select the 'Sync all sectors' button and click 'Yes'.



Return to the Assembly subitem, where you'll see the project contents starting to update. This process may take several minutes, so feel free to grab a cup of coffee while you wait. You can refresh the display periodically to monitor progress.

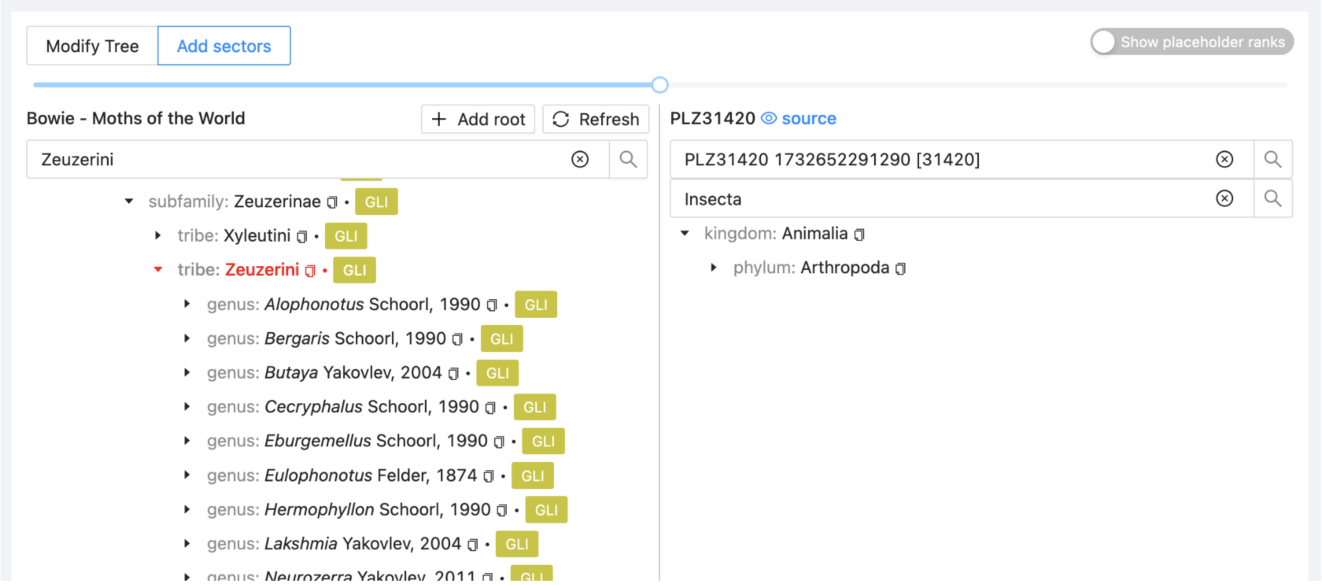


Alternatively, you may click on the double arrowed circle on the top right and see the Sync state.

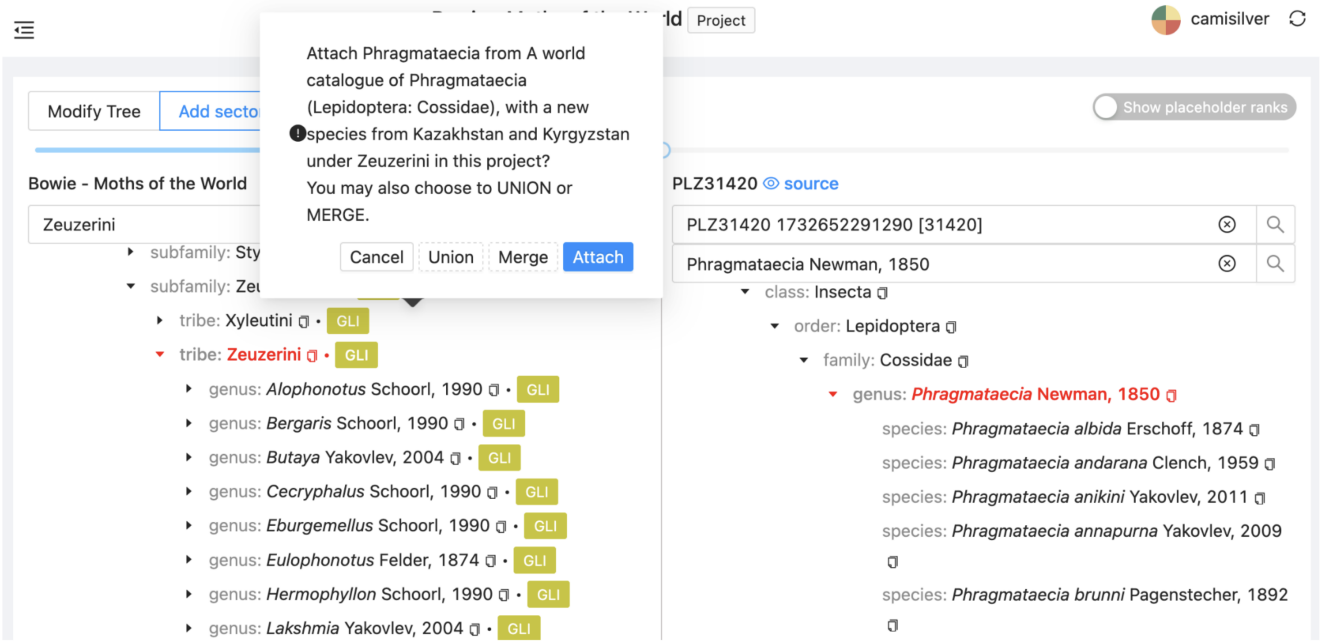


Now let's Attach a new sector.

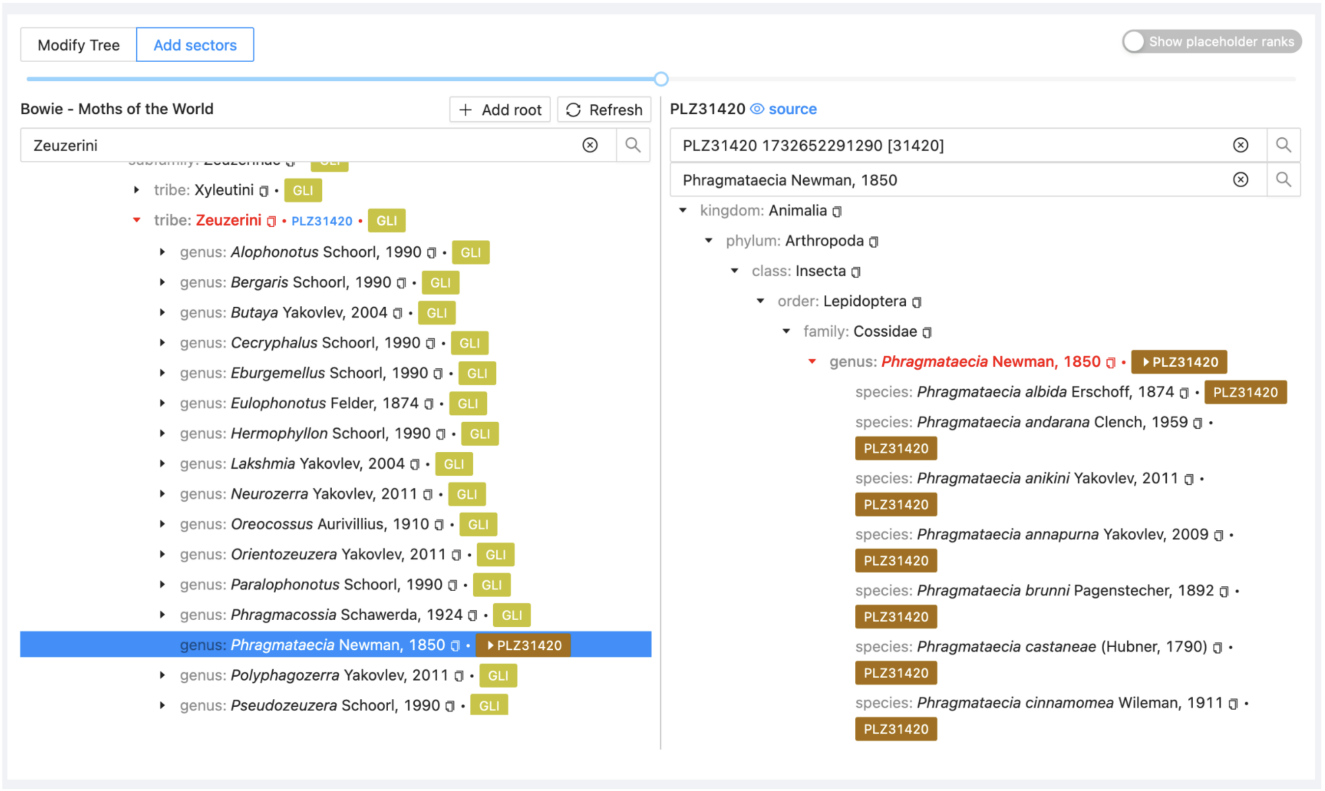
In the Assembly > Add sectors view, search on the left side for the tribe Zeuzerini and on the right search for Phragmataecia and select dataset PLZ31420 (PHR).



Find the genus Phragmataecia from PHR and drag it onto Zeuzerini in the project. Select Attach to insert the genus as a new sector.



A new sector (with a new colour) will appear on the left. The PHR (alias is PLZ31420). The sector has not yet been synchronised so it shows no child taxa.



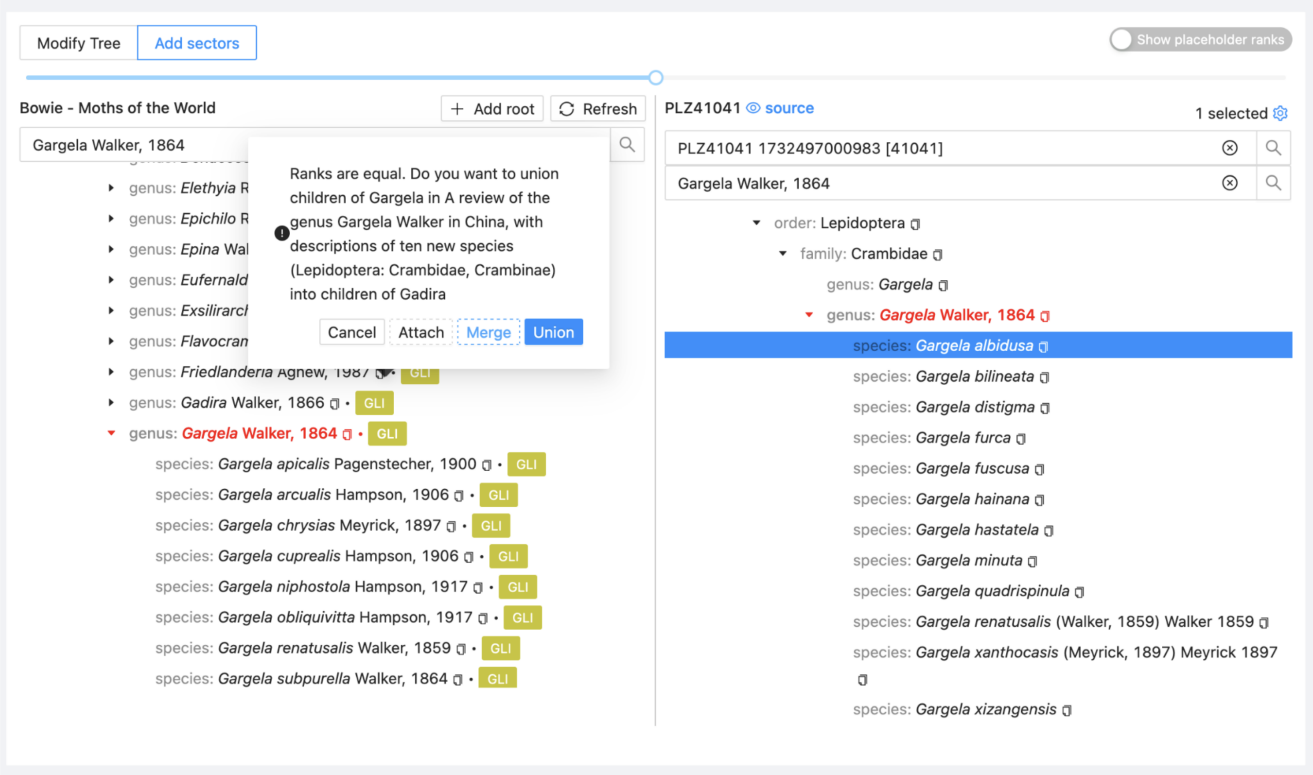
3.2.4. Merge

With this option to add sectors you can combine information from two or more sources. Here you will merge the species from GLI and those from the checklist derived from GAR, the paper “A review of the genus Gargela Walker in China…” (PLZ41041) that are not included in GLI.

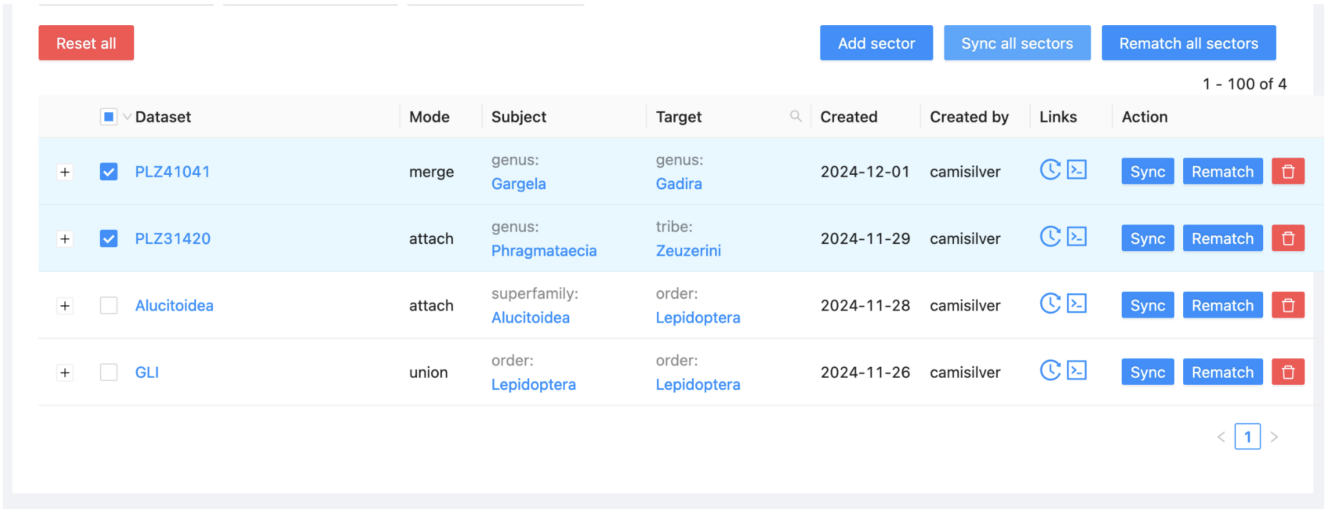
Go again to the Assembly > Add sectors view, search on the left side for the genus Gargela and on the right search for Gargela and select the dataset PLZ41041 (GAR).

Drag the genus Gargela from GAR onto the genus Gargela in the project view. Select Merge. It adds all

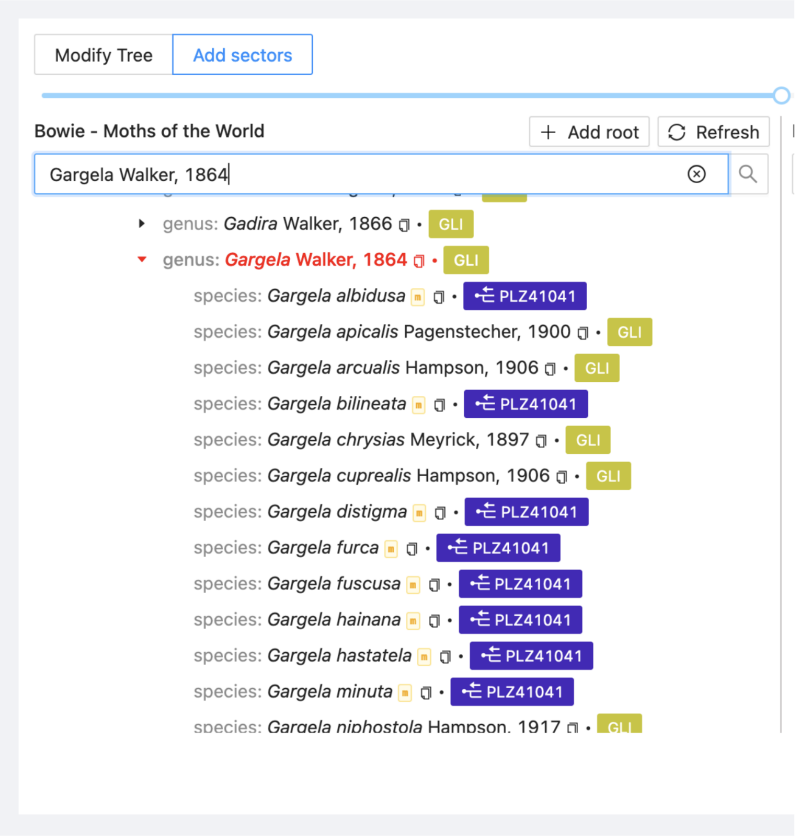
taxa from the source dataset to the project if there is not already an entry for them.



Go to the Sectors view and select Sync for the two new sectors (Phragmataecia and Gargela).



Following the synchronisation, the additional taxa will appear in the project view under Assembly.

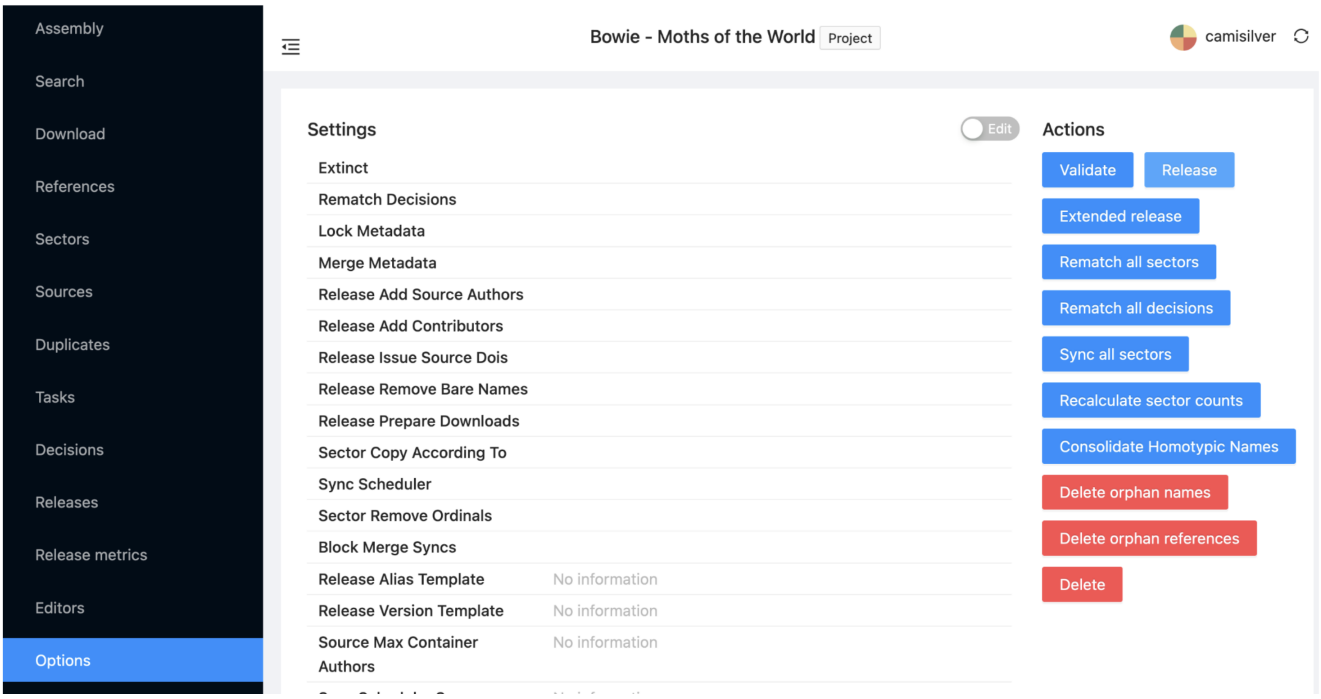


You can identify the merged taxa by a yellow square with the letter “m”.

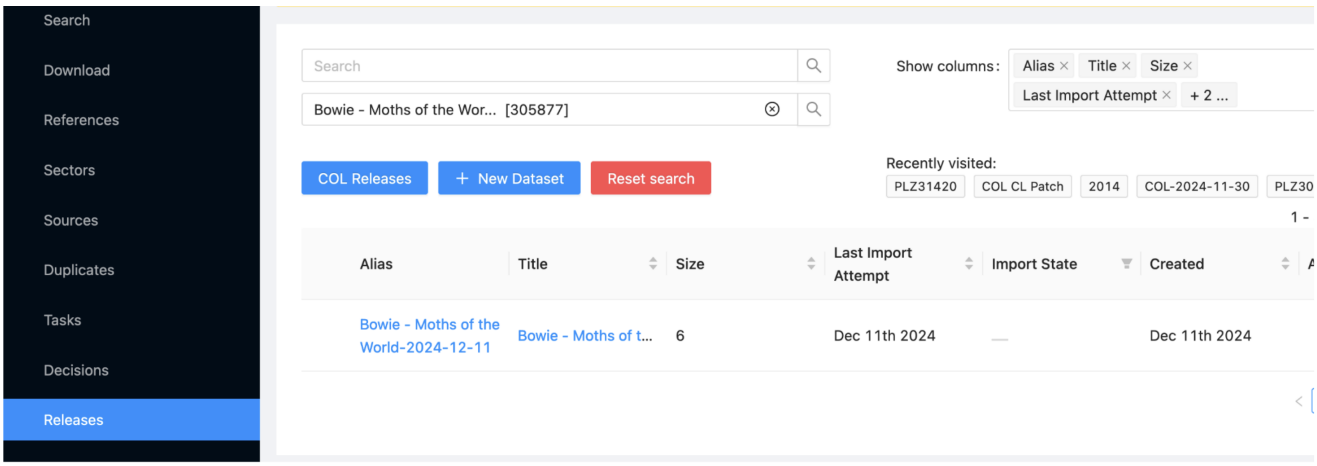
4. Release

To create a dataset that is a static snapshot of the assembly, you will need to make a ‘Release’. This will also generate stable identifiers for each taxon. The dataset generated will have the same characteristics and metadata as the project and will be available in the platform as public or private (depending on the definition of the project).

Go to the Options item on the left menu, and the Click on 'Release', the second action at the right. According to the project size it can take several minutes up to hours for the Release to be read.

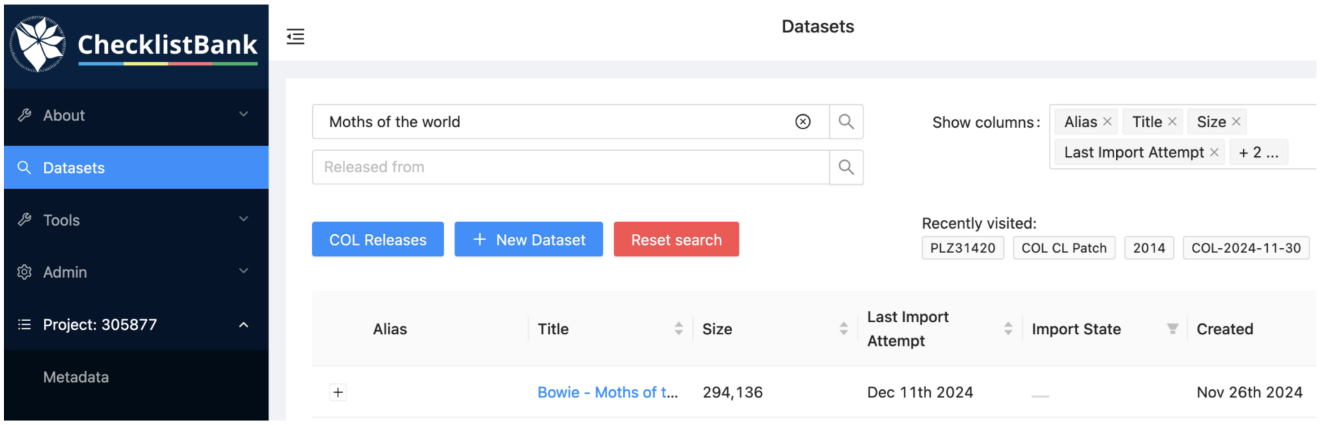


Releases are listed in the Releases item on the left menu.

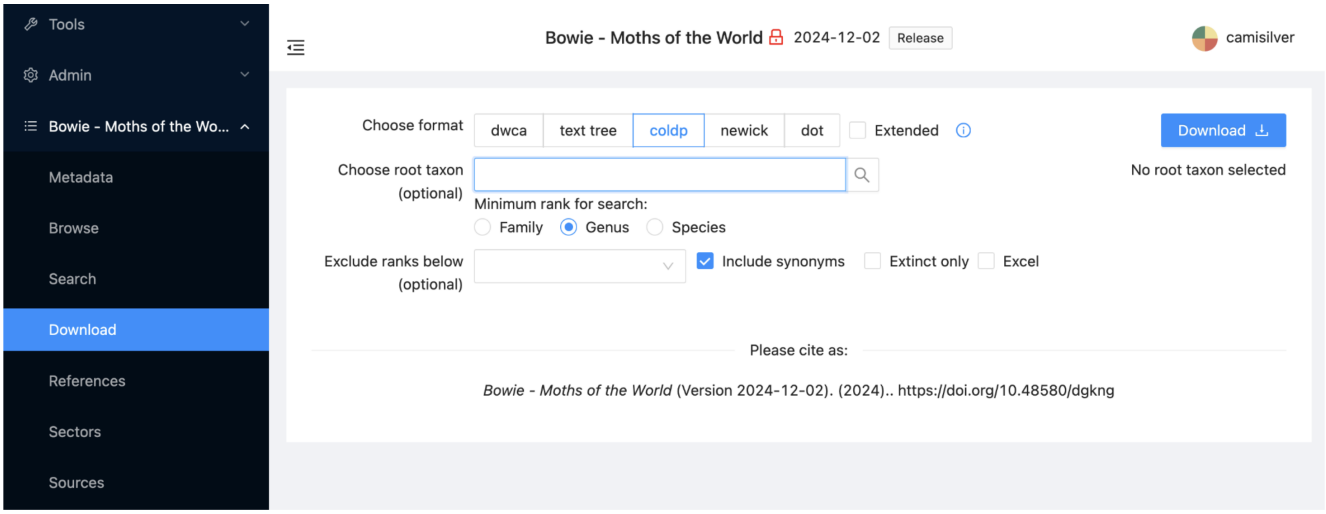


5. Download

Return to the main Datasets subitem on the left menu and search for your dataset. Once releases have been completed, you will find both the project dataset and its releases in the search results. Either can be selected for the following steps, although (as noted above) only the released versions will have stable identifiers if you plan to reference taxa by URI.

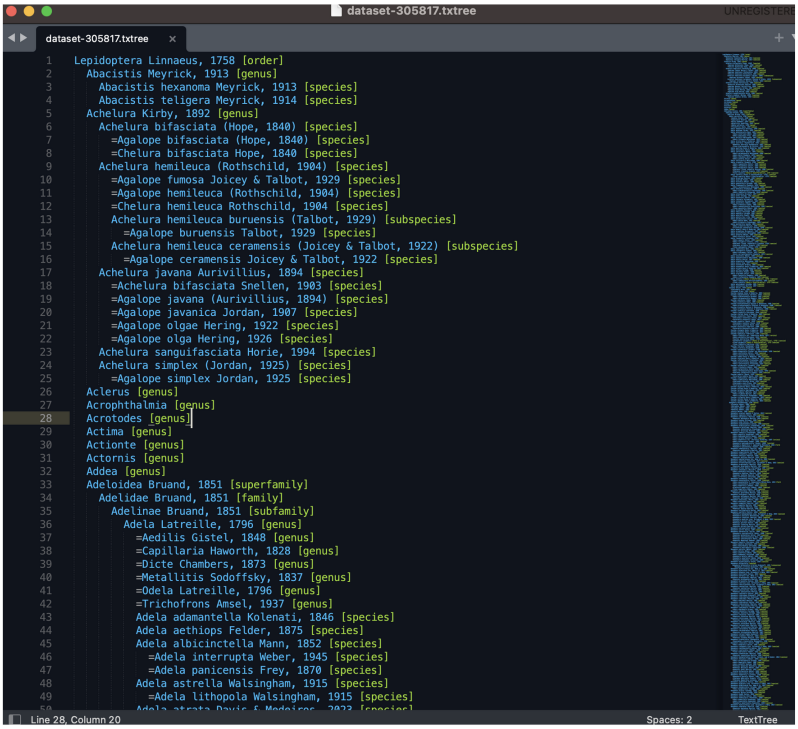


Select the Download subitem. This allows the dataset to be downloaded in a variety of formats:



- The coldp (Catalogue of Life Data Package) format is a rich data model that includes the most complete view of the contents of each dataset (see the CoLDP specification).
- The text tree format is a simple text view of the names, classification and synonyms. You will download this for your project dataset.

For this excercise apply any desired filters then select "text tree" and click on 'Download'. The resulting ZIP file contains the metadata (as a YAML file) and the tree representation of the dataset (as a TXT file).



The view shown includes the merged genus Gargela.

Colophon

Suggested citation

Hobern D., Plata C., Hernández D., Bánki O., Döring M., Stjernegaard T., & Miller J (2024). ChecklistBank Project Tutorial. A tool to build and curate checklists. version 1. https://doi.org/10.35035/ce-har0-r370.

Authors

Donald Hobern, Camila Plata, Diana Hernández, Olaf Bánki, Markus Döring, Thomas Stjernegaard Jeppesen and Joe Miller.

Contributors

This tutorial was developed to document the project functionality of ChecklistBank to support registered users interested in generating a new checklist based on the datasets available in the platform. Catalogue of Life acknowledges the European Commission funded project Biodiversity Community Integrated Knowledge Library (BiCIKL) (grant number 101007492), and support from the Netherlands Biodiversity Information Facility (nlbif2021.021).

Licence

Persistent URI

https://doi.org/10.35035/ce-har0-r370.

Document control

First edition, December 2024