



Taxonomic Editor

User manual:

Structured descriptions & Descriptive Datasets

Version: 29 January 2020

Contents

1	Introduction.....	1
2	Prepare for working with Descriptive Data Sets.....	1
2.1	Importing the Additivity Ontology.....	2
3	Create the Descriptive Data Set.....	2
3.1	Assigning taxa to the working set.....	3
	In order to assign the taxa to the Descriptive Dataset, you can simply drag and drop the taxon/taxa from the backbone in the Taxon Navigator into the Taxon filter field.....	3
4	Basics: Characters.....	4
4.1	Creating new terms for structure, property or state.....	5
4.2	Creating Term Trees.....	6
4.3	Assembling characters in the Character Editor.....	7
5	Assign the character tree to the working set.....	8
6	Specimen and Matrix.....	8
6.1	Import of specimens.....	9
6.2	Customize the table/matrix view.....	9

6.3	Enter data.....	9
6.3.1	New data.....	9
6.3.2	Literature data and default descriptions.....	10
6.4	Aggregate the datasets.....	10


1 Introduction

In a DFG-funded project (i.e. „Additivity Project“) an extension for the Taxonomic Editor has been developed to dynamically store, edit and analyse structured descriptive data. The workflow uses different perspectives and views to select taxa and terms and enter descriptive data. The software has been designed following the typical workflow of a taxon- and specimen-wise acquisition and processing of descriptive datasets. In the coming sections, the different components will be introduced and explained.

Central to the application is the so called Descriptive Dataset. The Descriptive Dataset contains the relevant taxa and the characters and states chosen to describe and distinguish them. Different filter allow restrictions with a regional focus or a limitation to certain taxonomic ranks of interest. After compiling the Descriptive Dataset, a tabular representation of the Descriptive Dataset can be opened (Character Matrix) and specimen as well as descriptive data from the literature can be associated, their individual data can be entered, processed and automatically aggregated to different ranks based on request. Finally, textual descriptions and polytomous keys can be generated and published via the Editors portal function.

In the interest of comparability, reusability and interoperability, all terms and vocabularies used to describe the taxa should come from a controlled vocabulary (ontologies), the scientific community has agreed on, is using and is constantly expanding and refining. The Editor is now providing several thousand terms for organismal structures, their properties and possible states. An internal *TermSearch* eases the location of certain terms within the extensive vocabularies. Missing terms can be search for and imported from several source-ontologies and vocabularies via an implemented Term Import using an own query form.

2 Prepare for working with Descriptive Data Sets

The Taxonomic Editor provides a specific perspective for working with Descriptive Datasets. To switch the Editor layout to this perspective, click on the Open Perspective symbol  on the top right and “*Choose Descriptive Data Set*” then click OK.

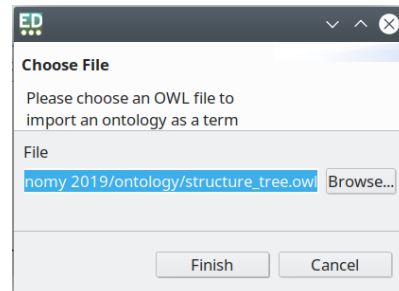
Alternatively you can open the Descriptive Data Set Navigator and related views by opening them individually via the main menu: *View > Descriptive Dataset*.

2.1 Importing the Additivity Ontology

The Additivity Ontology provides structure, property and state vocabularies and Term Trees (for an explanation of these see chapter below). Since a CDM data base initially has none such terms it is required to import these.

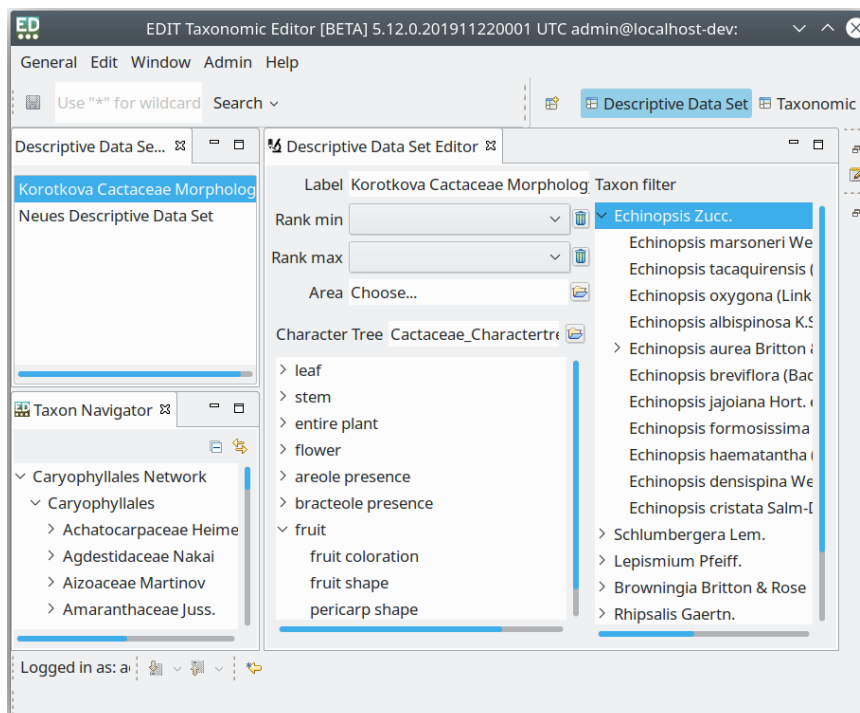
The terms are available to be downloaded from the EDIT developer Wiki: https://dev.e-taxonomy.eu/redmine/projects/edit/wiki/Additivity_Ontology . You will need the OWL files named *structure_tree.owl*, *property_tree.owl*, *states.owl* from this page.

Once downloaded these files open the OWL import dialog via the main menu: *General > Import > Descriptive Data > OWL import*. Choose the first of the downloaded files and click *Finish*. The import may take several minutes. Once the import is complete you can continue importing the other OWL files.



3 Create the Descriptive Data Set

To create a new Descriptive Data Set (for definition see introduction) right click in the Descriptive Data Set Navigator and choose *New Descriptive Data Set*. Double click to open the new created data set to open it in the *Descriptive Data Set Editor*.



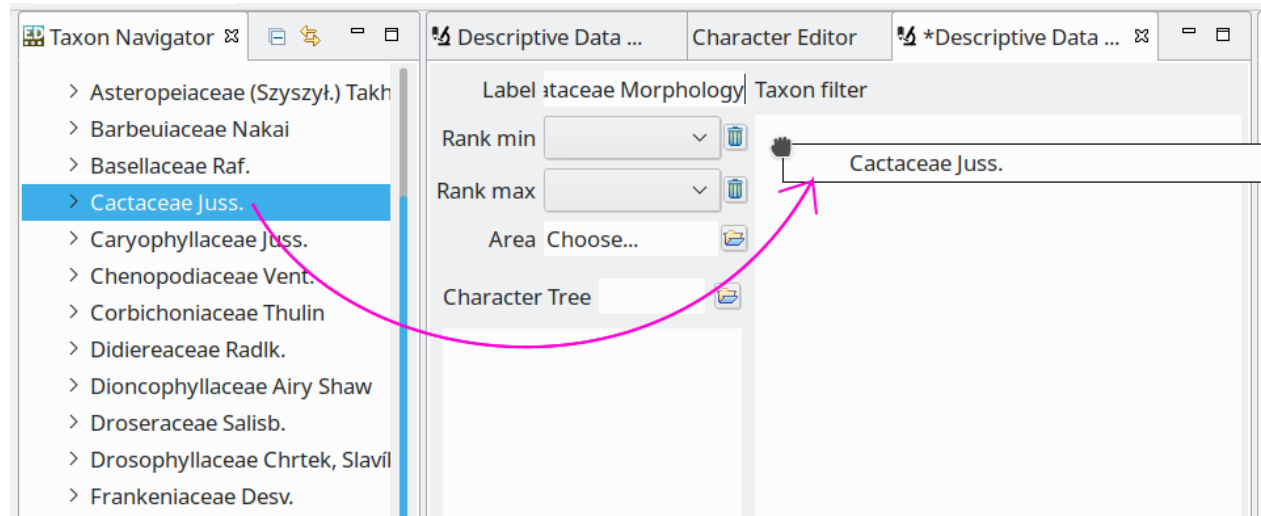
Label the newly created *New descriptive data set* by clicking on *Label*. Save modification by clicking the save button or use the shortcut *Strg+S*.

It is advisable to already limit the taxonomic extent of the final working set. This can be done here in advance by choosing a min/max rank of the taxa included. Additionally or alternatively a distributional limitation can be made.

Descriptive Dataset Navigator and Editor

3.1 Assigning taxa to the working set

In order to assign the taxa to the Descriptive Dataset, you can simply drag and drop the taxon/taxa from the backbone in the Taxon Navigator into the Taxon filter field.



What is missing now are the characters you want to record and analyse in the taxonomic group under examination. These characters must ultimately be provided in the form of a so-called character tree. How a character tree is created and how the characters it comprises are assembled, will be explained and illustrated in the following sections. Adding the Character Tree to the Descriptive Dataset is explained in chapter below.

4 Basics: Characters

Characters are composed of structures and properties. For example, the character “leaf colour” consists of the structure “leaf” and the property “colour”. The final character (leaf colour) can have multiple states (green, red, brown...). Hence, characters ultimately consist of three components: a **structure** with a **property** that has a **state**. For the sake of comparability and reusability, these components must be denominated with terms that are part of a controlled, globally (online) available, structured vocabulary in the form of a glossary or ontology. Once the Additivity Ontology has been imported (see chapter above), the Editor provides a wealth of terms that are pre-ordered and will cover most projects with

regards to plant morphology. Additional terms can be added to these extensive vocabularies via a Gfbio-Termimport to secure that they originate from a source that is in-line with the before mentioned demands and that they are properly stored with appropriate links and identifiers.

For reasons of clarity and to reflect natural relationship of terms and the things they describe, it is often advisable to give large vocabularies a hierarchical structure. This structure might correspond to the complex bauplan of an organism (morphological structures) or might only be a simple ordering system to provide a container for a certain group of terms that somehow belong together (e.g. state terms for e.g. colour).

The terms available in the Editor are initially pooled as flat lists in so-called TermTypes. Due to the large number of terms, structure- and state-terms are organised in vocabularies, the (much fewer) properties are only a list without any further classification. Whereas the states have been imported ordered in >50 predefined vocabularies, the structures have been organized in four artificial provisional vocabularies in order to enhance their handling (findability) when they are combined to individual TermTrees that subsequently are used to assemble characters.

A TermTree is a subset of terms from the three large TermTypes that has to be assembled for each structure and property needed to form the characters in the current project. The final list of characters is also a TermTree. A TermTree allows displaying the term relations in a hierarchical representation (e.g. structure and substructure – flower and petal). Additionally, it allows the repeated use of the same term, e.g. “leaf” as a repeatedly used substructure of other subordinate structures (e.g. stem and branch) but using essentially the same term that is uniquely represented in the Term vocabularies.

Finding existing terms

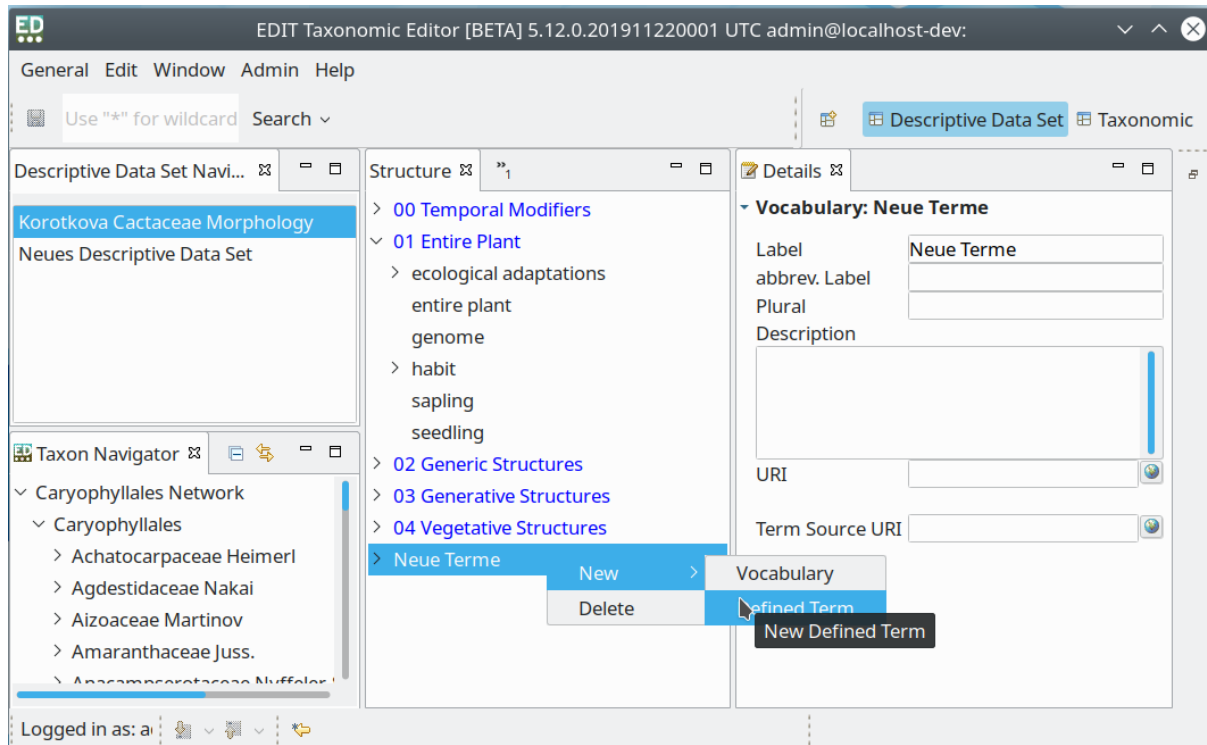
To easily find suitable terms for building the character required to describe a specific plant feature the Taxonomic Editor provides the *Term Search* which can be opened via the main menu: Window > View > Termsearch. The search result shows all terms containing the search string and shows in which vocabulary (TermType) this term can be found.

4.1 Creating new terms for structure, property or state

Before creating new terms it is highly recommended to search the database for this term (See 4.1). In almost all cases a suitable predefined term can be found in the vocabularies already contained in the cdm database (see chapter aboveabove). The database provides ca. 850 structure terms, 82 property terms and the state vocabulary already comprises more than 8500 state terms in 57 vocabularies. If nevertheless a certain term is missing, it can be added at any time.

To add missing terms (e.g. terms describing a structure, property or a state) to a TermType, open *Window -> Terms -> Others -> Structure* (or another Term Type)

The *Structure Editor* opens and you can either add/edit a term to an existing vocabulary or create either of it new by right-click in the *Term Editor: New-> Vocabulary*. You can label the vocabulary in the *Details View* and add a description, source, etc. Right-click on the vocabulary itself allows to add defined terms (right-click *New-> Defined Term*) or another (sub-)vocabulary now.



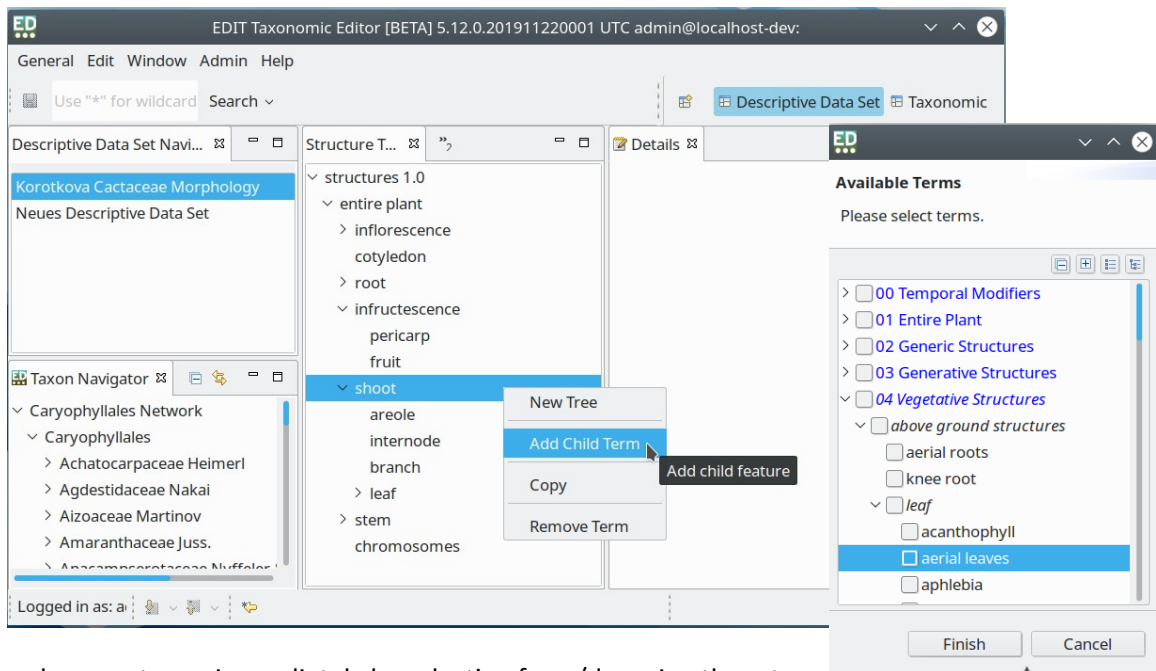
Name the newly created 'Untitled' Term in the *Details View* on the right side and save. It is advisable to add a definition and a source for the new term in the Details View, although it is not mandatory.

4.2 Creating Term Trees

As explained above, all terms needed in order to assemble characters must be preselected and (can optionally be) pre-ordered in a Term Tree. A Term Tree is a selection of terms from the vast vocabularies needed in your current project to describe the morphological features you want to examine in the taxonomic group of interest. These Terms can be ordered in a hierarchical structure at your discretion and practicability. Open the Feature Tree Editor, *Window->Term Editor-> Term Tree Editor->* e.g.

structure. The *Structure Tree Editor* opens and you can either use the existing Structure Tree expand/reduce it, or you can build your own Structure Tree from the scratch by: right-click->*New*. A “New Structure Tree” opens, you can change the label in the Details View in the field “Title Cache”. In order to add terms to the tree, you have two options:

1. Use: right-click->*Add Term* to choose from the structure vocabulary by checking the terms of interest, click ->*Finish* and the term will be added as a child element under your Structure Tree. Additional terms can then be added accordingly.
2. Alternatively, you can open a second window showing the structure vocabulary (*Window->Terms->Others->Structure*), expand it to the term you are looking for and drag and drop the term into your structure tree.



You can order your terms immediately by selecting from/dropping them to the correct parent structure. This can also be done at any time in the Tree Editor using drag and drop. Besides ordering terms hierarchically as parent and child terms (i.e. structure and substructure) you can also give your tree an (additional) artificial order in the form of a list. This allows a clear arrangement, especially when the tree gets bigger and allows the restriction of the term hierarchy to a certain level of depth by ordering structures into “groups” rather than forcefully maintaining the natural relations of structure and substructure that might be too complex to be reflected throughout.

Once your structure tree is created you have to do the same for the properties you want examine. Open the Term Tree Editor, *Window->Term Editor-> Term Tree Editor->Property*. The Property Tree Editor opens. *Right-click-> New*, a New Property Tree opens, change the label and add the terms you need using either the dialog (*right-click->Add Term*) or open the property vocabulary in a second window (*Window->Terms->Others->Property*) and drag and drop the terms into your Property Tree.

4.3 Assembling characters in the Character Editor

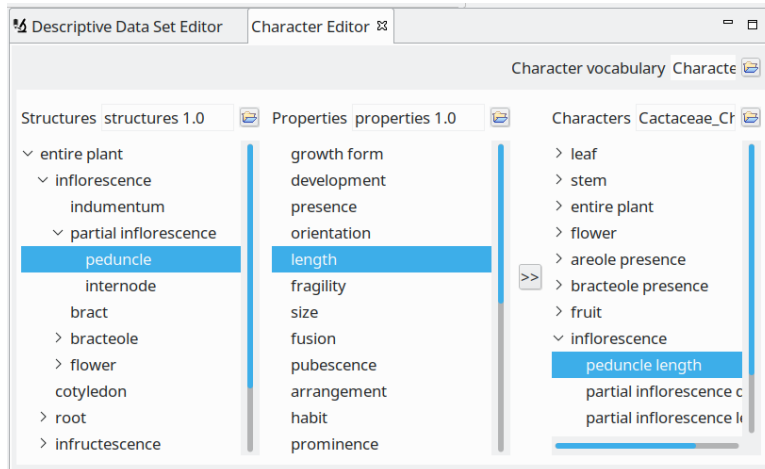
With the *Character Editor* structure and property terms can be combined to assemble a character. The characters themselves are organized in a Character Tree.

The states which are applicable are also part of the character definition. A state can be a defined term (categorical character) like in the case of named colours (red, blue) or can be a numerical value of a measure (quantitative character) in which case it is crucial to specify the unit of the measure. States and other properties of the characters can be edited in the *Details View*.

Open *Window->Terms-> Character Editor*

A (previously created) Term Tree with structures and one with properties must be chosen. By click on the folder icon next to Structures/Properties a dialog will open. By entering * in the upper input field, all Term Trees appear, click and select and press OK, to complete the process. The same has to be done for the Character Tree, either chose an existing one or create a new one using the dialog that opens upon clicking the folder icon. Additionally a Character Vocabulary must be chosen in the selection in the upper right. This vocabulary is necessary to store the characters assembled and their relations since these cannot be stored in the original vocabularies.

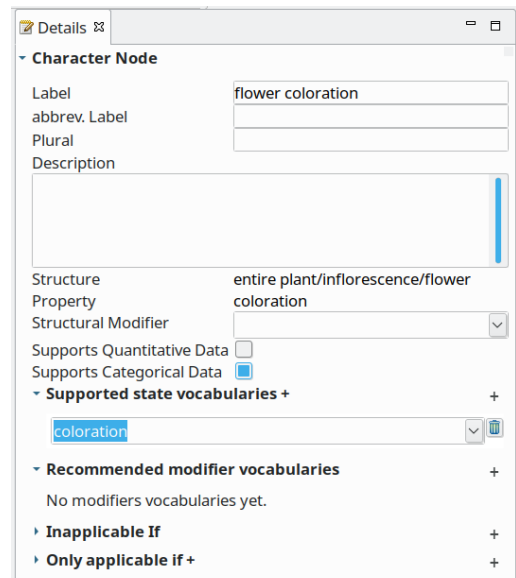
Combine Structures and Properties to Characters by clicking on one of the terms and press the >> Button to create.



To speed up the process, multiple selections of properties are possible by holding ctrl.

The newly created character appears at the right. The characters can now be related with each other using drag and drop. You can either drop one character on another one to specify that one is a subordinate character of the other. You can also sort them within the flat list by dropping them between two characters (black line appears indicating the right position to drop).

Finally, possible state vocabularies must be assigned to each character. Click on the character and on the right side a details view opens. You can either chose whether the character supports categorical data (terms) data or quantitative data (values). For categorical characters you can chose one or several state vocabularies. The state



vocabularies thereby largely correspond to the properties, sometimes there is a conceptual overlap, thus more than one state vocabulary can be selected to have all state terms available when entering the data. If the character is a numerical character, you can select a unit and the statistical values you want to record (e.g. min.-max. or exact value etc.).

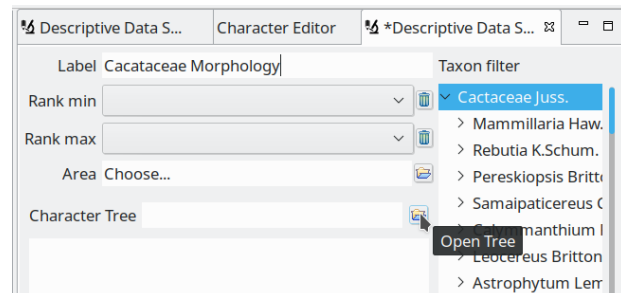
Note: Once a character tree has been built, (re-)ordering the characters and alter the state settings is also possible in the Term Tree Editor. Therefore open the Term Tree Editor, *Window->Term Editor->Term Tree Editor->Character*

5 Assign the character tree to the working set

Once the character tree has been built, the Descriptive Dataset can be completed and the data entry can get started.

Open the *Descriptive Dataset Editor* by right-click on the descriptive dataset in the *Descriptive Dataset Navigator*, selecting *Open in-> Descriptive Dataset Editor*. Now click on the folder symbol next to the box "Character tree". A dialog opens with a search field. In this input field you either enter the first letter of the created Character Tree that you want to use in your Dataset or you enter * to see all available trees.

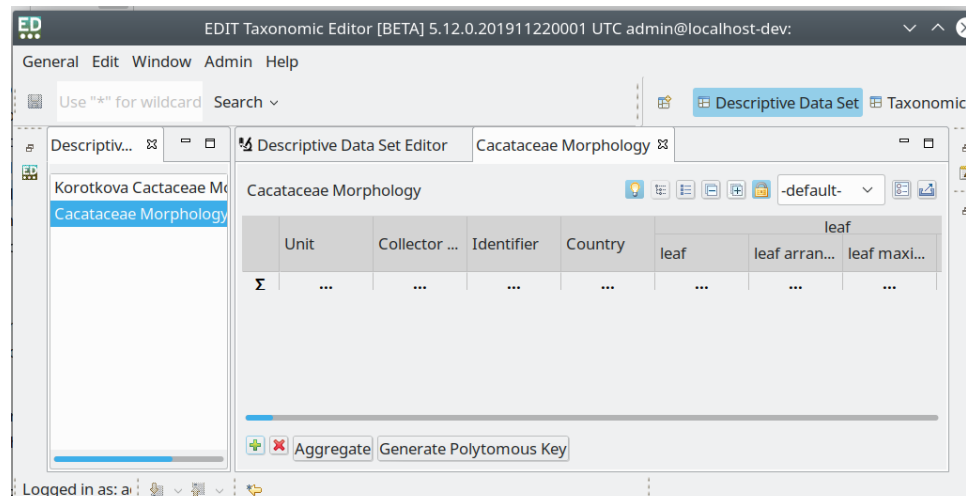
Click-> *OK*-> save, this character tree has been added to your dataset which is now completed.




Adding of the taxa to the Descriptive Workingset is explained in chapter

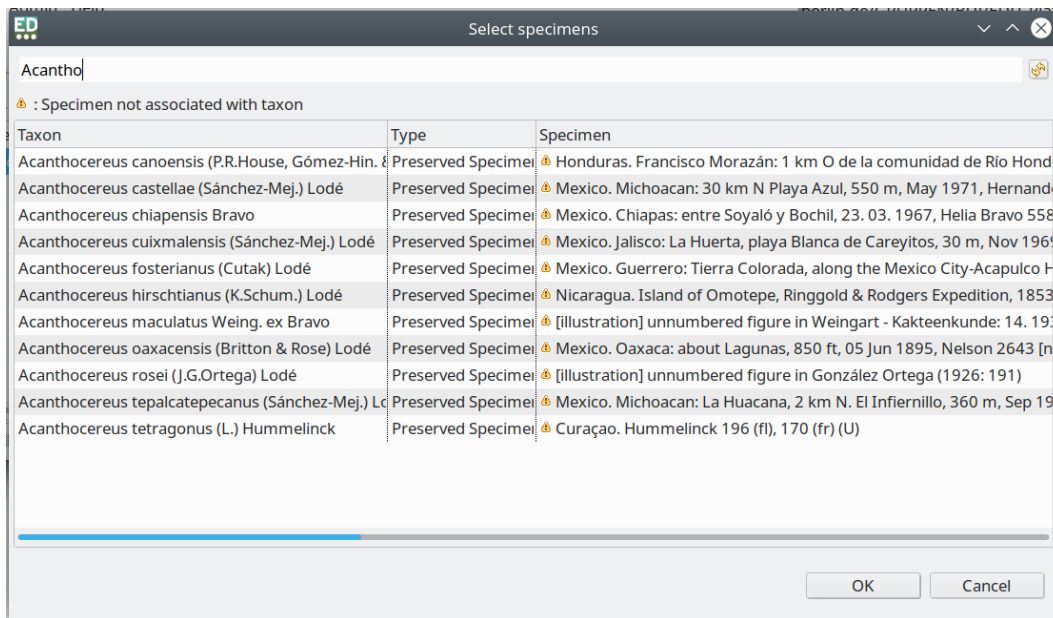
6 Specimen and Matrix

You can now start to work on the specimen and record the data according to your characters. Open the Descriptive Dataset Navigator -> right-click on the Descriptive Dataset -> Open in -> Character Matrix. The matrix will appear blank with the column headers showing the characters in the transposed order (vertical to horizontal) of the character tree. The hierarchy established in the character tree is transformed to headers and nested sub-headers in the table, thus providing blocks of character-complexes to improve clarity and ease data entry.



6.1 Import of specimens

To begin, you need to import the specimens into the matrix. Press the  button on the bottom left. A dialog named “Select specimens” opens providing a searchable list of all specimens found in the database that match your preselected taxonomic range. On top of the specimen list a text field is provided which will filter the list by the text entered into it.



Select the specimens by clicking on the respective lines (multiple selection by using ctrl.) and press ->OK. The specimens are now imported into your descriptive dataset and will appear ordered line-by-line according to their current determination. Depending on the extend of your taxonomic pre-selection, every rank will hierarchically be represented by one line, with the specimens allocated to them, depending on the specificity of their actual determination. Overall, ranks are ordered alphabetically. Metadata (e.g. Collector, Collection number, etc.) will automatically be allocated to the corresponding cells in the columns at the beginning of the table. To delete specimens, select the whole line (click on the number) and click the X button at the bottom.

6.2 Customize the table/matrix view

Depending on the number of characters examined, the matrix can become very large. To enhance usability a number of personal configurations can be made.

- use the lock symbol to freeze the informative columns at the beginning of the table
- use right click -> *hide/show columns* to reduce the screen-view to the (currently) relevant columns

- combine both functions to reduce the freezed info-columns to your preferences

Finally, all personal configurations can be saved in the configurations view in the toolbar in the upper right. Stored configurations can then be chosen from the drop down in the following sessions. This allows different users to work with their preferred view, returning to the default configuration can be done at any time.

Also the list of taxa and specimen can be very large and might be confusing when entering data specimen- or taxonwise. It is advisable to collapse all but the lines you are currently working on. The collapse/hide all buttons in the toolbar might be helpful in this context.

6.3 Enter data

6.3.1 New data

Entering data into the matrix is very intuitiv. Just click on the respective cell and different dialogs open depending on the character.

For categorical states a drop-down opens showing a list of all possible character states contained in the state vocabulary(ies) previously assigned to the character. These often very extensive lists of states are ordered alphabetically and can be searched using auto-suggest. To use this, just type the first letters of the desired state-term and the list will automatically jump to the best match. Hence, the more letters entered, the better the suggestion will be, until a direct hit. Use the checkboxes for multiple selection, click on the term itself to change from on term to another.

If the state is a numerical one, a dialog opens showing all fields to enter the statistical values previously selected when assembling the character. Enter the values in the respective field(s), click OK or press enter and they'll appear in the matrix. Use the bin symbol in the dialog to delete erroneous entries.

6.3.2 Literature data and default descriptions

If you want to compare your newly generated data (taxa or characters) with previously published datasets, you can enter the data from other (published) sources for every taxon. In order to do so, select the taxon in the matrix (left click) and then right click->*Create literature description*. A new line appears below the selected taxon. In the details view on the right you can enter a source and additional information regarding the origin of the data.

Enter the data as described above. The same workflow applies for a default description if desired.

Default data might be helpful if you know that all datasets (specimen) for this taxon have the same state for certain characters but you want to include them nevertheless (e.g. discriminative between taxa).

If you want to delete any of these additional datasets, select the whole line (click on the line number) and press the X button at the bottom.

6.4 Aggregate the datasets

If you have entered the data for all (or some) taxa and specimen you can at any time press the *Aggregate* button. Aggregations are automatically done at every taxonomic level. The aggregation appears as a new line directly under the respective taxon.

