



User manual, appendix:
Preparing a Caryophyllales family
contribution for the
World Flora Online Initiative's
taxonomic backbone

Version: 22 July 2019

Contents

1	WFO Data Definition for the WFO taxonomic backbone.....	1
2	General process for treatments prepared in the EDIT Platform.....	1
3	Preparing the DwC-A file.....	1
	Testing the DwC-A file.....	2
	A note on excluded names and taxa.....	3
4	The DwC-A meta.xml file.....	4



1 WFO Data Definition for the WFO taxonomic backbone

See the WFO Guidelines for Taxonomic Backbone Contributors:

<https://www.dropbox.com/s/78h02zhnct7ikqh/WFOGuidelinesForTaxonomicBackboneContributorsV.%202.06.pdf?dl=0>

2 General process for treatments prepared in the EDIT Platform

In the WFO database, the so-called Content data (descriptions, distributions, common names, etc.) are linked via a WFO Identifier to a taxon name in the so-called Taxonomic Backbone. The WFO Identifier (wfo-ID) is an arbitrary number with the prefix “wfo-“, assigned by the WFO-consortium to (eventually) all botanical names of vascular plants and mosses.

Management of the Taxonomic Backbone follows rules different from Content management. The taxonomic integrity and the updating of the backbone is done by replacing the existing “taxonomic slice” (e.g. a family) in the backbone. All names already in use in the WFO-Backbone **must** be included in the update so that all existing wfo-IDs will again found in the replaced backbone segment.

One of the first steps in the preparation of a backbone contribution is thus obtaining the backbone slice existing in the WFO database, with all its names and accompanying wfo-IDs, to make sure that these are treated. New names will receive a new wfo-ID. The process of assigning and/or reuse of wfo-IDs will be accompanied by the BDI-Section at the BGBM.

Once the treatment is complete, a Darwin Core Archive file is prepared in accordance with the WFO Guidelines, quality checked, and sent to the WFO Gatekeeper for further checking and eventual integration in the WFO database.

3 Preparing the DwC-A file

The following specifications refer to the Caryophyllales species level database in the EDIT Platform (caryophyllales_spp) but can be used for other groups as well. Examples are taken from the Nepenthaceae treatment.

NB: wfo-IDs have to be already assigned to all names in the family for the following procedure to work.

You need to have MS-Access installed (procedure tested for Access 2013 and 2016).

Create a local folder for your data processing. If you have it already, delete all content in the folder.

Open the TaxEditor and, in the TaxonNavigator, right click on the taxonomic group (usually a family) you want to export. Choose *Export* and *CDM-light (csv)*. As the folder for the export, use your local folder.

Click *Next* and make sure, that the checkbox *Create Zip File* is not checked.

Click *Finish* and wait for the export to conclude (progress shown in lower right corner of TaxEditor).



From Y:\BDI\EDIT-Plattform_Latest Versions\CDM-light-2-WFO-Backbone-DwC-A, copy the following files into the local folder where the exported *.csv files are located:

- CDM2WFOtb*.accdb
- meta.xml
- schema.ini
- WFO.bat

Double-click on the CDM2WFObackbone*.accdb in your local folder

If you get a security warning, click “Enable Content”

Click on *Import CDM-light ...*

Enter the family when prompted

Wait until the program finished

Close the database.

As a result, there should be a directory WFO-DwC-<familyname> in your target directory, with the tables classification.txt and references.txt, and the meta.xml file.

Use the Windows function (right click on the directory) to compress these three files into a .zip file, which is to be submitted to WFO.

Testing the DwC-A file

The following instructions are given in the WFO General Guidelines for Contributors (v. 2.07) to test the resulting DwC-A file:

The DwCA file can be tested with the GBIF DwC-A Validator service under <https://www.gbif.org/tools/data-validator>. Another tool provided by GBIF called the Darwin Core Archive Validator (<http://tools.gbif.org/dwca-validator/>) also serves to check for some of the most common problems when generating a DwCA file. However, please note that the non-DwC elements in the DwCA will be marked as errors, as well as the link (parentNameUsageID) of the family to the higher taxon ID (either because the ID is missing or because the taxon referred to - e.g. an Order name- is not present in your file). Also, the test will let some terms pass that are not accepted by WFO (e.g. “nom. nov.”, “nom. nud.” etc. as values for nomenclatural status) – please refer to the respective lists of permitted values in the Guidelines for Backbone providers.

There are some queries in the access database (numbered from 01 on) that serve to control the output with respect to the links within the Classification table (01-02) and with the References table (03) , and with respect to adherence to the WFO controlled vocabularies (04-06).



A note on excluded names and taxa

The WFO backbone specification includes a data field called *exclude* in the *classification.txt* table. In the WFO context, this field has a double purpose – any text in this field will exclude this name record from the WFO taxonomic backbone that is used in the WFO portal (while maintaining the wfo-IDs as exclusive to this record). The text in the field should point to the reason why the name was excluded, essentially for the WFO Gatekeeper to be able to retrace the underlying decision.

On the EDIT Platform's side, there are two possibilities to introduce an exclusion into the CDM2WFO backbone program. The name that should be excluded has to appear as a taxon node, i.e. you must be able to see it in the taxon navigator. The preferred way to exclude it is to mark it as such in the *Taxon node wizard* (right click at the node, *Open in*) and type in an *Excluded note*. N.B.: this text is mandatory. An alternative is to assemble names that should be excluded under pseudotaxa, i.e. a textual entry in the classification tree that indicates the reason for the exclusion. For example, in the Nepenthaceae dataset, under the Family, there is a pseudotaxon called "Nepenthes published artificial hybrids". To achieve exclusion of names contained under this category, you assign a dummy wfo-ID to the pseudotaxon name. This dummy ID has the format DUMMY_ followed by the text you want to appear in the exclude field of the WFO-output file. In the given example, that is DUMMY_Published_artificial_hybrid (the underlines will be replaced by space in the output).

Note that this should be an ID referring to the Name (not the Taxon), of the type *World Flora Online Plant Name Identifier*.



4 The DwC-A meta.xml file

```

<?xml version="1.0"?>
<archive xmlns="http://rs.tdwg.org/dwc/text">
  <core encoding="UTF-8" linesTerminatedBy="\n" fieldsTerminatedBy="," fieldsEnclosedBy="&quot;"
ignoreHeaderLines="0" rowType="http://rs.tdwg.org/dwc/terms/Taxon">
  <files>
    <location>classification.txt</location>
  </files>
  <id index="0"/>
  <field index="0" term="http://rs.tdwg.org/dwc/terms/taxonID"/>
  <field index="1" term="http://rs.tdwg.org/dwc/terms/scientificNameID"/>
  <field index="2" term="http://rs.worldfloraonline/terms/localID"/>
  <field index="3" term="http://rs.tdwg.org/dwc/terms/scientificName"/>
  <field index="4" term="http://rs.tdwg.org/dwc/terms/taxonRank"/>
  <field index="5" term="http://rs.tdwg.org/dwc/terms/parentNameUsageID"/>
  <field index="6" term="http://rs.tdwg.org/dwc/terms/scientificNameAuthorship"/>
  <field index="7" term="http://rs.tdwg.org/dwc/terms/family"/>
  <field index="8" term="http://rs.emonocots.org/terms/subfamily"/>
  <field index="9" term="http://rs.emonocots.org/terms/tribe"/>
  <field index="10" term="http://rs.emonocots.org/terms/subtribe"/>
  <field index="11" term="http://rs.tdwg.org/dwc/terms/genus"/>
  <field index="12" term="http://rs.tdwg.org/dwc/terms/subgenus"/>
  <field index="13" term="http://rs.tdwg.org/dwc/terms/specificEpithet"/>
  <field index="14" term="http://rs.tdwg.org/dwc/terms/infraspecificEpithet"/>
  <field index="15" term="http://rs.tdwg.org/dwc/terms/verbatimTaxonRank"/>
  <field index="16" term="http://rs.tdwg.org/dwc/terms/nomenclaturalStatus"/>
  <field index="17" term="http://rs.tdwg.org/dwc/terms/namePublishedIn"/>
  <field index="18" term="http://rs.tdwg.org/dwc/terms/namePublishedInID"/>
  <field index="19" term="http://rs.tdwg.org/dwc/terms/taxonomicStatus"/>
  <field index="20" term="http://rs.tdwg.org/dwc/terms/acceptedNameUsageID"/>
  <field index="21" term="http://rs.tdwg.org/dwc/terms/originalNameUsageID"/>
  <field index="22" term="http://rs.tdwg.org/dwc/terms/nameAccordingToID"/>
  <field index="23" term="http://rs.tdwg.org/dwc/terms/taxonRemarks"/>
  <field index="24" term="http://purl.org/dc/terms/created"/>
  <field index="25" term="http://purl.org/dc/terms/modified"/>
  <field index="26" term="http://purl.org/dc/terms/references" default="http://www.worldfloraonline.org" />
  <field term="http://rs.tdwg.org/dwc/terms/kingdom" default="Plantae"/>
  <field term="http://rs.tdwg.org/dwc/terms/nomenclaturalCode" default="ICNAFP" />
  <field term="http://purl.org/dc/terms/rightsHolder" default="World Flora Online Consortium"/>
  <field term="http://purl.org/dc/terms/license" default="https://creativecommons.org/publicdomain/zero/1.0"/>
  <field term="http://purl.org/dc/terms/rights" default="CC0 1.0 Universal (CC0 1.0)." />
</core>
</archive>

```