

CDM import process notes

ver. 1.0

Thomas Hamann

Copyright: Document copyright © Thomas Hamann/Naturalis Biodiversity Center 2015-2016. This document is licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) license.

Table of Contents

CDM import process notes	1
General notes.....	1
Possible problems causing import errors	1
Checking the data portal.....	3

General notes

The purpose of this document is to provide some notes on the Naturalis side of the CDM import process. Once all obvious corrections have been made to a FlorML file and it validates in XML Spy, it is sent off to Andreas Müller at the Botanical Garden Botanical Museum Berlin, who handle the hosting of the CDM databases and the accompanying Data Portals of each of the three Naturalis flora series (Flora Malesiana, Flore du Gabon, Flora of the Guianas). Using scripts he developed, he tries to run the import process. If the import process is successful, the data contained in the FlorML file is imported into the CDM database, and will show up on the data portal. In that case, have a look at “Checking the data portal”. If the import process produces errors, read the next section.

(For details regarding the scripts for importing FlorML files into the CDM database, please contact Andreas Müller)

Possible problems causing import errors

During the import process, the scripts try to match each taxon mentioned in the keys to the accepted name in the nomenclature of the actual taxon treatments. A failure to match a taxon name in a key to an actual taxon can have several reasons. Each of these is shortly discussed below, including solutions:

- Non-standard rank abbreviation. The import scripts expect that the rank abbreviations used for sub- and suprageneric ranks (e.g. subspecies, section)

follow the currently used standards. In older floras this often is not the case. The solution here is to replace non-standard abbreviations with the standard ones.

- Non-standard name format, e.g. a variety of a subspecies. These need to *exactly* match what is in between the fullName tags, including eventual author names. This also applies to hybrids, which also should use the proper mark-up to explicitly identify them as hybrids.
- Mismatch between the taxon name given in the key and the one given at the actual taxon treatment because one of the two names is wrong. Possible reasons are (a) an OCR error in either the key or at the taxon treatment, (b) the taxon name changed following the application of an erratum, or (c) the author of the treatment made an error.
In the case of (a) and cases of (c) where the error was a typing error, correct the error. In the case of (b) and cases of (c) where the problem is that two different names are given, replace the name in the key by the correct one. Note that sometimes the error is that the taxon number is wrong; in those cases fix the taxon number.
- Mismatch cause by additional text present in key. A taxon name in a key is followed by text providing some comment on that taxon name, e.g. “p.p.” or a comment between brackets. Put such comments in keynotes tags after the toTaxon tags.
- Mismatch caused by reference to different taxon with the same rank as the taxon the key belongs to, or the same rank as the taxa in the key but a different genus, e.g. a key under Rhododendron referencing Vaccinium. There currently is no real solution for this, so you have to request a manual match.
- Mismatch caused by reference to taxon present under other taxon of the same rank as rank of taxon key is located under, e.g. a key belonging to a section that references species that were already mentioned under another section earlier and actually belong to the latter section. There currently is no real solution for this, so you have to request a manual match.
- Footnote references within the accepted name nomenclature. There currently is no real solution for this, so you have to request a manual match.
- Taxon numbers that are wrong or missing. Fix the number.
- Actual taxon treatment does not exist. Sometimes, keys reference taxa that are (a) not actually discussed within a particular flora volume, (b) taxa only mentioned in the key but outside the geographical scope of the flora, or (c) cultivated taxa only mentioned in the key.

The solution here is to insert a dummy taxon for each taxa that is only mentioned in the key. For taxa for which only a name is given, give it a nomenclature section with only that name and a description with the text from the key. For taxa that are outside the geographical scope, use `inGeoScope="false"` and also include the distribution information. For cultivated taxa, use `class="cultivated"`.

- Actual taxon treatment cannot be found due to some other reason, but taxon treatment is actually present in the flora volume. Check whether the `<taxon>` tags are present for the taxon treatment involved, as well as the immediately preceding and following taxon treatment. Check whether the accepted names of these three taxa are correctly marked as such. Check whether any synonyms of these three taxa are incorrectly marked up as accepted names.

The import scripts also try to link references to couplets within a key to the corresponding couplets. The following error can occur:

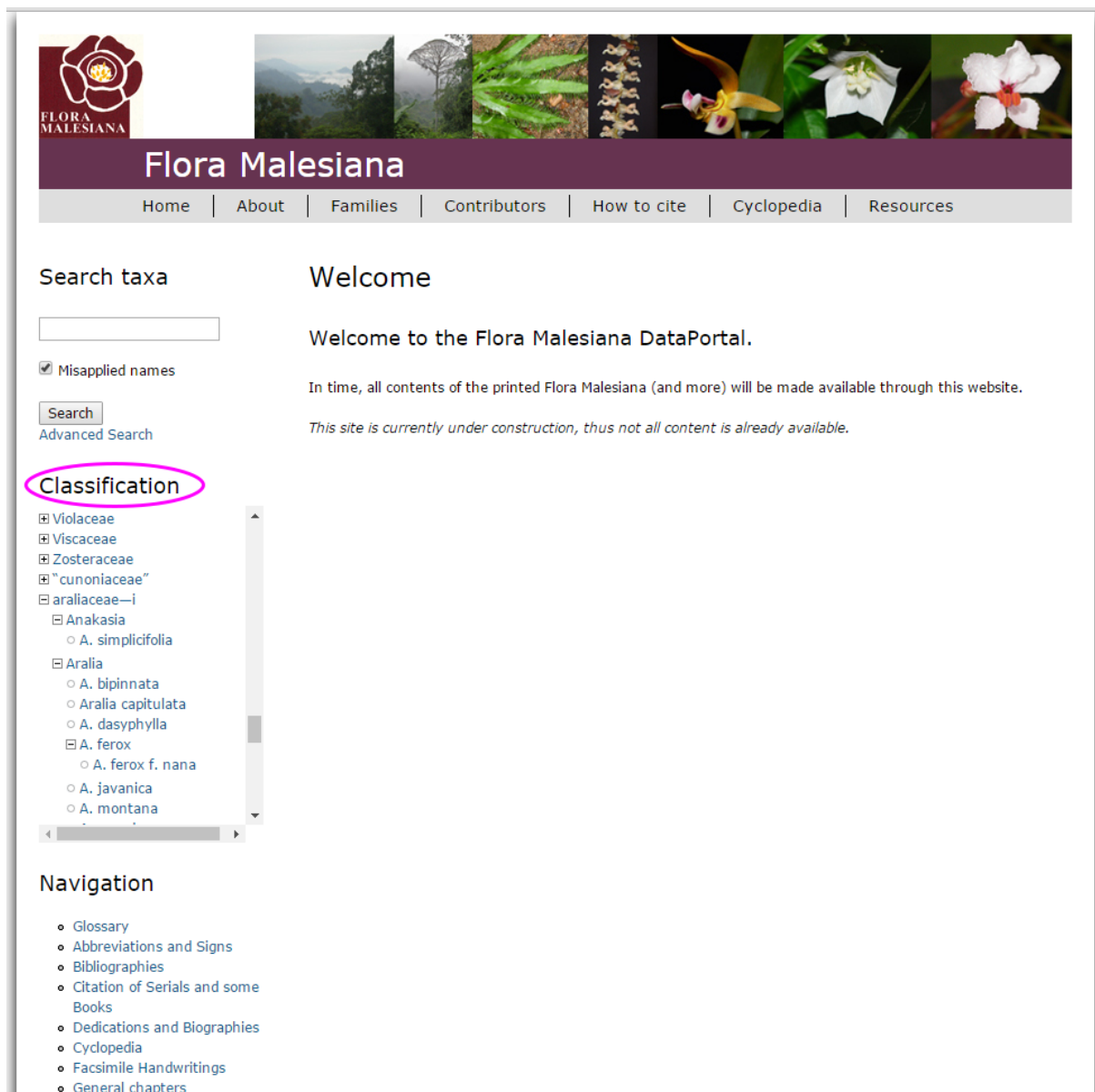
- Couplet matching errors. These usually are caused by incorrectly split up couplets, wrong numbers in couplets, or wrongly ordered couplets, and mean a couplet is referenced by a question but cannot be found. The number preceding the error hash is the couplet number the script is looking for. Search for all cases of such a reference, and check whether the referenced couplet exists. Then correct the problem.

Many of these problems can be taken care of during the proofreading stage, but especially in larger keys it is easy to miss something.

Once these errors have been fixed, the import should succeed.

Checking the data portal

Once the data has been imported into the CDM and made available in the data portal, it may be useful to check whether all taxa are located where they ought to be located according to the paper volumes, i.e. whether they are placed under the correct parent taxa. It is fairly simple to do this by expanding the taxa shown under "Classification" in the Data Portal (see Figure 1, next page). However, it is important to note that excluded taxa currently end up under the taxon from which they are excluded from without it clearly being indicated in the "Classification" that they are excluded (a case of still being in development). If a taxon is incorrectly located, this can be an indication that there is an error in the mark-up with regards to the taxon rank.



Flora Malesiana

Home | About | Families | Contributors | How to cite | Cyclopedia | Resources

Search taxa

☒ Misapplied names

[Advanced Search](#)

Classification

- Violaceae
- Viscaceae
- Zosteraceae
- "Cunoniaceae"
- Araliaceae—i
 - Anakasia
 - A. simplicifolia
 - Aralia
 - A. bipinnata
 - Aralia capitulata
 - A. dasyphylla
 - A. ferox
 - A. ferox f. nana
 - A. javanica
 - A. montana

Navigation

- Glossary
- Abbreviations and Signs
- Bibliographies
- Citation of Serials and some Books
- Dedications and Biographies
- Cyclopedia
- Facsimile Handwritings
- General chapters

Figure 1: Flora Malesiana data portal with "Classification" circled, showing several expanded taxa.