User manual, appendix:

Processing CDM-light output with MS-Access and MS-Word

W. Berendsohn

Version: 21 march 2024

Contents

1   Preparing the environment

You need to have MS-Office installed (procedures tested for Office 2013, 2016, and 2019).

Create a local folder for processing your taxonomic treatment (if you are working with different EDIT Platform databases, you should create at least one folder each). If you have it already, delete all content in the folder with the exeption of the file *facts2use.csv*.

Open the TaxEditor and, in the TaxonNavigator, right click on the taxonomic group you want to export. Choose *Export* and *CDM-light (csv)* [1]. Use your local folder as the target for the output.

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 the TaxEditor).

Copy the following files into the local folder where the exported *.csv files are located:

- CDM2Word.accdb (the MS-Access database used for further processing)
- schema.ini (a configuration file ensuring that the data can be correctly imported – truncated text and erroneous diacritic characters result when this file is not present)
- settings.csv (a configuration file that determines if, e.g., types or sec.-references are included in the output – see section 3 below).
- A facts2use.csv file, if not present (see section 4 below for further information)

   and, in case you plan to further process the data with Word

- CDM2Word-Format-yyyy-mm-dd[2].dot (the MS-Word template used for further processing)
- CDM2Word-Style.dotx (the MS-Word template defining the paragraph styles used)

Double-click on the CDM2Word.accdb in your local folder - If you get a security warning, click "Enable Content"

Choose if you want to create the processed output tables in the database that are needed for further processing in word. Note that depending on the size of your data, this may not be possible (e.g., do not try to use the entire Euro+Med Plantbase with >100.000 names and >30.000 taxa).

Click on *Import CDM-light …*

Wait until the program finished (a message will pop up).

If you want to continue to process the output in Word, close the database and proceed to chapter 2. If you want to manipulate data in Access, proceed to section 5 below.

---

[1] See *Manual-EDIT-Platform-Appendix-CDM-Light-*.pdf* for a detailed description of the CDM-light format.
[2] The template version is indicated by the date in the filename

2   Generating a document using MS-Word

2.1   How to

Close all open Word documents.

Doubleclick on CDM2Word-Format-yyyy-mm-dd[3].dot.

If you get a security warning, click "Enable Content"

Enter the name of the Family (and Genus) your output is about.

Wait until the program finished (a message will pop up).

The result can be found in the subdirectory "output", the filename contains the Family name entered in the beginning and the date and time of the creation of the file.

2.2   Background (work in progress)

The CDM2Word-Format*.dot template

The CDM2Word-Style.dotx template

Data cleaning

In the Word program, numerous replacements are done to eliminate repeated full stops, spaces in the wrong place etc.

---

[3] The template version is indicated by the date in the filename

3   The settings.csv file

The settings.csv file is a table with values that control certain aspects of the output of the programs. Some are used by the Access database, some by Word, and some by both. The table contains two columns: The parameter (abreviated function of the entry) and the value (true or false, i.e. -1 or 0). An explanation of the parameters (their effect if set to true, i.e. -1) is given in the following table.

DoIndSciNm	Access: Write a concordance RTF File that is used by Word to create an index of scientific names (including WFO-IDs). Word: Use the RTF file to create an index of scientific names. Note that abbreviated Names in text notes etc. are not indexed.
DoIndComNm	Access: Write a concordance RTF File that is used by Word to create an index of common names.) Word: Use the RTF file to create an index of scientific names.
DoClean	Word: Text cleaning functions activated.
Debug	Word: Activates code used for debugging
InclSupraspTypes	Word: Include supraspecific types
InclSpeciesTypes	Word: Include species (and infraspecies) types
InclFactRef	Word: Include references for facts
InclSecRef	Word: Include the secundum reference
InclSynSecRef	Word: Include the synonym secundum reference
InclProtologURI	n/a
DoEnglish / DoSpanish	Word: Determines the output of propts for type statements.
InclMisapplied	Word: Include misapplied names
GenericTypeNewLine	Word: Type species of the genus not put on the same line as the name and nom. reference
DoWordOutput	Access: Process data and create tables that are necessary for the Word program. Deactivating this may allow to import very large amounts of data that otherwise surpass the size limits of Access.
DoURIIDList	Access: Write an RTF file (SciNameIDURI.rtf) with all scientific names, their WFO-ID (optional) and their protologue URI (optional). This file can be opened by word and used independently.
InclWFOID	Access: Include WFO-ID in SciNameIDURI.rtf
InclURI	Access: Include protologue URI in SciNameIDURI.rtf
InclIDURIList	Word: Include the SciNameIDURI file as an Appendix in the output file.

4   The Facts2use.csv file

This is a table with 3 columns that determines, which fact categories are included in the output, in which sequence, and what prompt should be inserted before the fact's value.

Include the fact categories using the english term, if it exists. Of course, only fact categories used in your project can generate output.

Example (from eFloraMex):

FactCategory	FactSequence	FactPrompt
Citation	1	Cita
Notes	3	Notas
Description	5	Descripción
Forma de crecimiento	6	Forma de crecimiento
CondensedDistribution	7	Distribución
Distribution (General)	9	Distribución general
Distribution	11	Distribución
Tipo de ambiente	12	Tipo de ambiente
Common Name	13	Nombre(s) comun(es)
Uses	15	Usos
Conservation	17	Conservación
CITES 2020	18	CITES 2020
IUCN 2021-2	19	IUCN 2021-2
NOM-059-SEMARNAT 2010	20	NOM-059-SEMARNAT 2010

5   Further processing of CDM-light data in Access

5.1   Warning: Updates replace the database!

The CDM-light text tables are present in the CDM2Word database under the same names as in the CDM-light export (without the .csv file extension). Whenever you start the import procedure, these tables are replaced with the up-to-date versions you exported from the Taxonomic Editor. This database contains a large number of additional tables, modules, queries etc. that are used for the document generation process with MS-Word. The functionality is constantly improved and (hopefully not that constantly) debugged. This means that you will have to replace the entire database from time to time.

For quick lookups and ad-hoc queries, you can use the CDM2Word database directly. However, if you want to introduce your own objects, e.g. queries for statistics etc. that are specific to your data and that you want to use frequently, you have to use a separate database that contains the tables from the CDM2Word database as a backend, as linked tables[4]. These can be used for queries etc. just like tables in the database. Updating the tables in CDM2Word automatically updates these tables in CDMFrontend, and replacing CDM2Word with a new version will not break the links. So you are free to define your own queries etc.

5.2   Predefined queries for data quality control

There are a number of more or less simple queries you can use to control your data. They are found in the selection query and the union query section under "All Access Objects" and they carry the prefix "DQ-" for data quality. Their name is rather self describing, but there may be some additional useful functionality. The queries come with a standard sorting, but you can change that freely in the tabular view. Some of the names are self-explanatory, for others please find a brief explanation below.

DQ-Names with Generic Reference
DQ-Names with Identifiers
DQ-Names with Type Information
DQ-Names without Ref Detail
DQ-Names without Ref Year
DQ-Names without WFO-ID
DQ-Synonyms partial pro-parte
DQ-Taxa without Sec Ref   DQ AppendedPhrase from Taxon and S...
DQ-URIs with Names, Title Sort   DQ-Unpublished Taxa and Synonyms

---

[4] This has certain advantages against using the .csv tables directly as linked tables. Apart from better performance, it avoids the problems caused by the string handling of MS-Access during the import – without specific measures, text strings will be truncated to 254 characters.

DQ-Misapplied Names with references

A list of names assigned to a taxon as misapplied names (under the synonym list). The sensu reference gives the source where the misapplication was made, the err-sec reference is the reference that has cleared that up (that said that the concept under the misapplied name belongs to the taxon).

DQ-Names with Generic Reference

A list of names that have a nomenclatural reference that has not been properly atomised (generic reference type)

DQ-Names with Identifiers

A list of all names with their name IDs and their type:

Acantholimon acerosum var. parvifolium Bokhari		
Acantholimon acerosum var. persicum Mobayen	100280759	Tropicos Name Identifier
Acantholimon acerosum var. persicum Mobayen	wfo-0000513147	WFO Name Identifier
Acantholimon acerosum var. pinardii (Boiss.) Mobayen	wfo-0000513148	WFO Name Identifier
Acantholimon acerosum var. pinardii (Boiss.) Mobayen	100280760	Tropicos Name Identifier
Acantholimon acmostegium Boiss. & Buhse	wfo-0000513150	WFO Name Identifier
Acantholimon acmostegium Boiss. & Buhse	685888-1	IPNI Name Identifier
Acantholimon aegaeum F.K.Mey.	932255-1	IPNI Name Identifier

DQ-Names with Type Information

A list of all names stating type related information (if any) from CDM-light fields (ProtologueTypeStatement, TypeSpecimens [concatenated for Word output), TypeStatements. In addion, a field HasParenthesis indicates names that cite a basionym (and thus should not have a type).

DQ-Names without Ref Details, … without Ref Year

List of names lacking Detail or Year with their Reference data items

DQ-Names without WFO-ID

DQ-Synonyms partial pro-parte

List of synonyms that have been flagged as partial or pro-parte in the database

DQ-Taxa without Sec Ref

DQ-Taxa Synonyms Unpublished

Taxa and Synonyms that are flagged as unpublished.

DQ-Taxa Synonyms Appended Phrase

List of taxa and synonyms that have an appended phrase

DQ-URIs with Names, Title Sort

A list of names with their nomenl. References and URIs of their original publication. Sorted by reference so as to identify works that have partly been covered, partly not.