

USER MANUAL

TEMPLATE ABCDIMPORT2DARWIN: ENCODING DATA IN XML-FILE (ABCD SCHEMA STRUCTURE)

Version		Author(s)	Date	Description
1.0	ABCDImport2DaRWIN _Entomo_v1_0.xlsm	MAdam	January 2014	Testing version of the template for import
1.1	ABCDImport2DaRWIN _Entomo_v1_1.xlsm	MAdam	March 2014	Production version of the template for import
1.2	ABCDImport2DaRWIN _Entomo.xlsm = "CLASSICAL" version	MAdam	August 2014	Small adaptations and debugging
1.2 bis	ABCDImport2DaRWIN _Entomo_taxonFullName.xlsm = "_taxonFullName" version	MAdam	December 2014	New version with simplified taxonomy

For information about ABCD - Access to Biological Collection Data: <http://wiki.tdwg.org/ABCD>

TABLE OF CONTENTS

December 2014 – Updates	3
Method 1 – Use of the revised template (_taxonFullName version) and of the template for taxonomy import	3
Method 2 – Use of the template released in September 2014	3
Comparison	4
September 2014 - Updates	5
User information	5
Template structure.....	10
Additional information to fill the template	11
1. Introduction	11
2. ID number	11
a. Column names for IDs	11
b. Duplicated IDs	12
3. Taxonomy	12
4. TAXONOMY-sheet (only in “classical” template).....	13
5. Dates	13
6. People name	13
7. Sampling code.....	13
8. Container storage	14
9. Pre-export checks	14
Technical information – ABCDschema TAB	16
1. Commands	16
a. Export group.....	16
b. Tools group.....	16
c. Checks group	17
2. Output.....	17
a. Name and extension.....	17
b. Structure.....	17
Glossary	18
XML and ABCD XSD schema.....	18
Visual Basic For Applications and macros	18
APPENDIX.....	19
1. General list of supported fields with expected format, description and example(s)	19
2. Link between old and new field names	20

DECEMBER 2014 – UPDATES

After extensive tests made within the IT, it was shown that the “Create missing taxon” tool could lead to unforeseen errors. It was therefore decided that the creation of taxonomy had to be made in a previous step to the specimens import. A template for importing of taxonomy has therefore been prepared.

This template for taxonomy comes in an excel format. Each taxonomic level corresponds to a column. An export in xml format produces an xml file that can be imported directly into DaRWIN via an import interface for taxonomy, replacing the "Create missing taxon" button that is not available anymore.

For more information about this template, you can also read the user manual for the taxonomy import template.

From now on, two methods are available for specimens import...

Method 1 – Use of the revised template (_taxonFullName version) and of the template for taxonomy import (*preferred method*)

The principle for this method follows the “DaRWIN good practice” that is to prepare your taxonomy previously to the import of specimens into DaRWIN.

How does it work ?

You have 2 templates at your disposal: the revised template for specimens

“ABCDImport2DaRWIN_General_taxonFullName.xslm” and a new template for importing taxonomy “TaxonomyImport.xslm”.

The template for importing taxonomy should be used to create the missing taxonomy corresponding to the specimens you wish to import. More information about this template is available in the dedicated user manual.

In the new specimens template, only the “taxonFullName” field remains for taxonomy. You have to give the taxon name/identification with the exact same spelling than what is present in the “name” field of the taxonomic entry in DaRWIN. Remember that if you wish to give the full taxonomy in the template, you can add as many columns as you wish. At the moment of the export, a popup message will appear, warning you that unrecognized columns will not be present in your export file...

Besides, a tool was also developed to check if an entry already exists in the Taxonomy Catalogue in DaRWIN. It is the “Check_taxonFullName.xslm » excel file. You need to enable macros and have an active internet connection to make it work properly. This tool will, through a custom ribbon and button, scan the DaRWIN database, to see if what is in the “taxonFullName” column already exists. You only have to copy-paste the column “taxonFullName” from the specimens template, to the “taxonFullName” column in the “taxonFullName” worksheet of the excel file, and click on “CheckTaxon” in the custom ribbon. This way, you can check before your specimens import, if the taxon name you used in the “taxonFullName” columns will be recognized. If no, you can prepare a taxonomy template for importing taxonomy or create the missing entries directly through the DaRWIN interface.

Method 2 – Use of the template released in September 2014

The template released in September 2014 is still functional. You can continue to work with it. Only the use of the “taxonFullName” column is slightly different.

What is different ?

When you import a file, during the specimens check, you will be asked to select the corresponding entry in the Taxonomy Catalogue in DaRWIN (or to create the missing entry) for specimens which have unrecognized taxon names. Only one correction by taxon and by file is necessary: the name will be corrected automatically in each line where the taxon name is the same.

Warning ! Keep in mind that the column “taxonFullName” is used for checking if the taxon name is already present in DaRWIN.

> For identification up to lower taxonomical levels (genus or below), the column “taxonFullName” is used for checking if the taxon name is already present in DaRWIN. It is therefore strongly advise to use the exact same taxon name in this field “taxonFullName” than what is present in DaRWIN. If this column is empty at the moment of the export, it will be automatically produced by concatenation of what is in the “genus”, “species”, « subspecies », « form_variety » and « author_year » columns. Risks of errors are therefore high...

> For identification to higher levels (above genus), the system check the name based on the lowest level referenced in the template.

When you fill many levels for taxonomy in the template, there is also a risk for unmatched if the name is spelled differently in DaRWIN or if the hierarchy is different in DaRWIN (one sublevel missing for example). We therefore advise you to specify only one or two higher levels corresponding exactly to what was encoded in DaRWIN and be careful with the name spelling. Example: only fill the “genus”, “species” and “taxonFullName” columns or, for an identification up to the genus level, only fill the direct parent, like the “familia” column.

Comparison

	Method 1 <i>ABCDImport2DaRWIN_General _taxonFullName.xlsm</i>	Method 2 <i>ABCDImport2DaRWIN_General.xlsm</i>
« taxonFullName » column	Only way to precise an identification; if empty, no identification... Can be a name for <u>any taxonomical level</u> .	Used to compare with the names in Taxonomy Catalogue in DaRWIN, for <u>taxon names from genus</u> (and levels below); if empty, automatically recomposed.
Higher levels and taxonomical hierarchy	No higher levels available in the xml ABCD exported file.	Higher levels, when specified, have to be spelled exactly the same way than in DaRWIN and follow the exact same hierarchy...
Conditions for immediate matching during specimens import	An entry in the Taxonomy Catalogue in DaRWIN has to exactly match with what is specified in the “taxonFullName” column of the specimens template. Warning! The matching is made for the character string specified in the “taxonFullName” column without special characters and lowercase (parenthesis, for example, are not taken into account for the matching).	For identification until genus level or below: > taxonFullName matching to a DaRWIN entry in the Taxonomy Catalogue, for the lower identification level that was specified in the specimens template > for each specified parents in the template, exact same spelling and hierarchy than in DaRWIN for the corresponding entry is obligatory > Advise: specify only the direct parent that exists in DaRWIN. For higher levels identification: > taxonFullName not useful > for each specified parents in the template, exact same spelling and hierarchy than in DaRWIN for the corresponding entry is obligatory > Advise: specify only the direct parent that exists in DaRWIN.
Additional tools	> “TaxonomyImport.xlsm” to prepare taxonomy before specimens import > “Check_taxonFullName.xlsm” to check if names in “taxonFullName” column are recognized	

SEPTEMBER 2014 - UPDATES

- New fields for specimen storage corresponding to the “Container” Widget in DaRWIn:
 - o Container
 - o Container Type
 - o Container Storage
 - o Sub Container
 - o Sub Container Type
 - o Sub Container Storage

The boxStorage and tubeStorage fields are not available anymore since there was a conflict with these new fields.

- New field: samplingCode
- Since the use of a dot (“.”) in specimen ID is frequent, several tests were made to see if an error could occur. It appears that the use of a dot surrounded by letters is not a problem. Example: INV.2367. Remember that it will be stored in the “code” field of the “Codes” widget and not split into a prefix and a code in the code widget. If you wish to add a prefix for one collection, the curator of the collection should edit the collection to explicitly give a default prefix to the code. A function was also added to the macro, so that some special characters are corrected. Nevertheless, it is still important to avoid special characters in Excel cells.
- Improvement of the export function in the macro.

USER INFORMATION

Template tested with Excel 2007.

1. Open the template document and save it with a custom name “[CUSTOMNAME].xlsm”.

If a Security Warning appears, then Enable macros (click options... and select “Enable this content”).

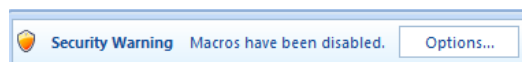


Figure 1 - Security Warning for macros

You can also access these options through the Excel Options > Trust Center ❶. Then, click on “Trust Center Settings” ❷. In the new window, in “Macro Settings” ❸, check “Enable all macros” ❹.

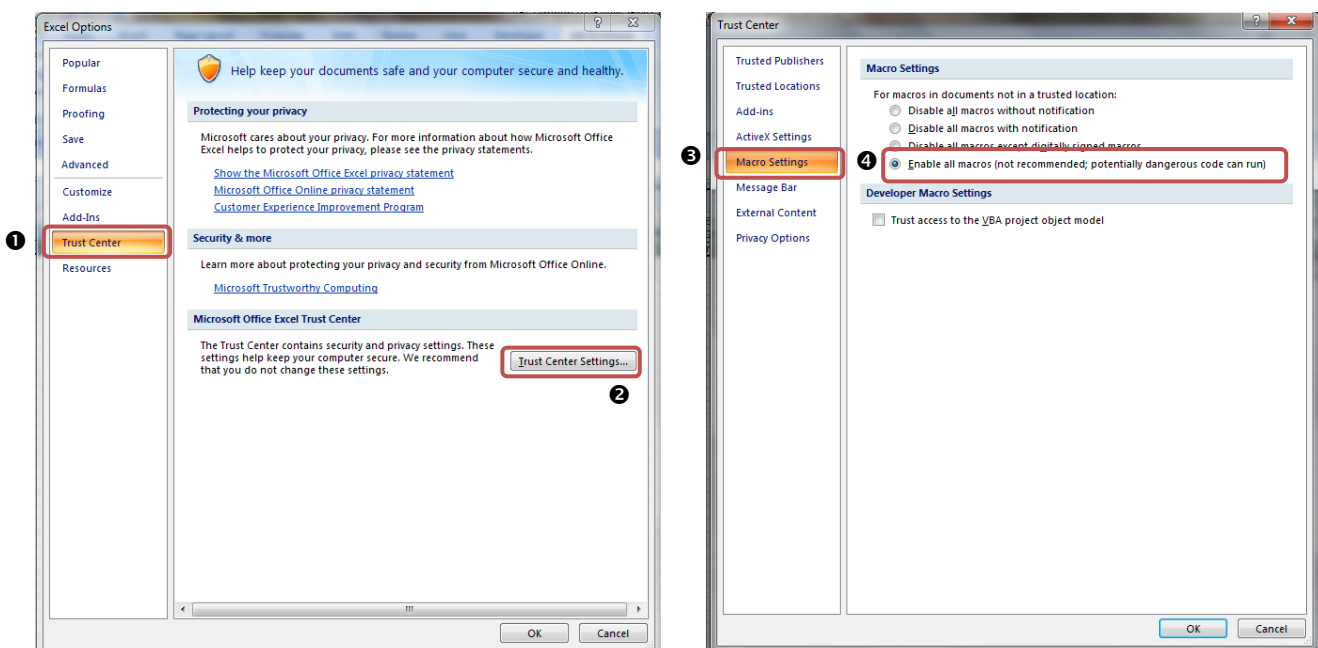


Figure 2 - Excel options: enable macros

2. Fill in the worksheets SPECIMEN and TAXONOMY

Information about the specimen is stored in a worksheet named SPECIMEN. In the “classical” template, information about the taxonomy (phylum, class, order and family) is stored in a worksheet named TAXONOMY.

3. Before export, some checks can be made

For more details about the tools available for checking before export, see Technical Information. The *Quick Practical Guide* below gives you some practical information on how to use of the template. It gives you some minimal advices to complete the template properly. Nevertheless, we strongly advise you to read this user manual completely.

4. Click on “Export2ABCD” to export the data

When running the Export2ABCD code, an XML-file following the ABCD schema structure is produced and this XML-file can be saved in a folder defined by the user. During the export, Excel is unavailable (the worksheet could disappear or turn into blue during the process). This may take several minutes, depending on the number of lines and the quantity of information.

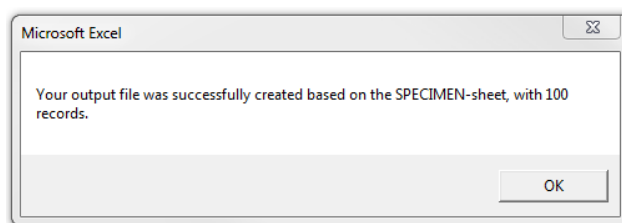


Figure 3 - Summary of your export

QUICK PRATICAL GUIDE

Template structure

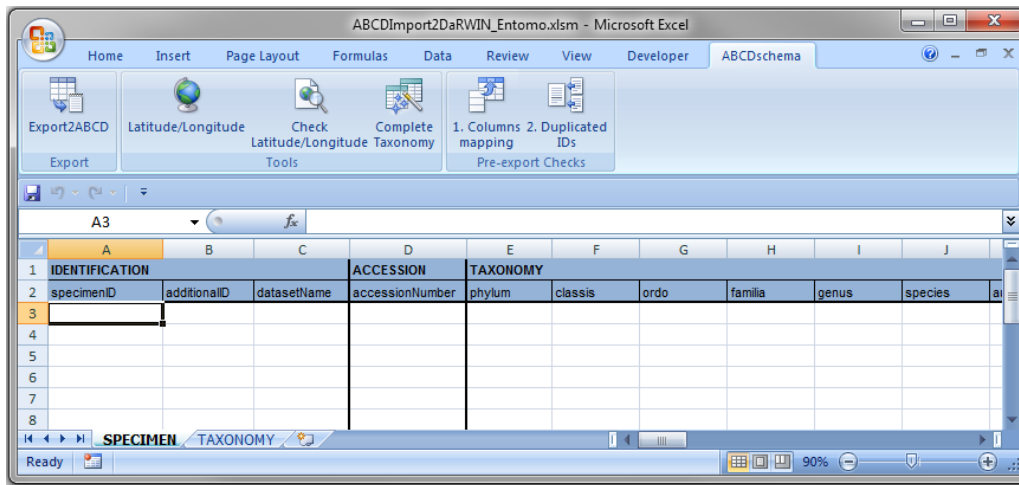


Figure 4 - Template structure

Verify the presence of:

- **A worksheet named 'SPECIMEN'** containing information about your records and, in the “classical” version of the template, a **worksheet named 'TAXONOMY'** containing information about taxonomical levels higher than the family name
- **A title for each column in the SPECIMEN-sheet**, written in the **second row** and using the exact same name and spelling as in the pre-established list of supported fields available in the appendix of this document. If this condition is not fulfilled, the information will not be exported to the XML ABCD formatted file. You can add as many columns as you wish, for internal purpose but the information they contain will not be exported to the XML ABCD formatted file.
- **In the “classical” version of the template: 4 columns in the TAXONOMY-sheet named 'phylum', 'classis', 'ordo' and 'familia' and written in the first row**
- **A column for IDs in the SPECIMEN-sheet**, named 'specimenID'
An ID is not required, but remember that links between specimens and hosts or other kind of units (e.g., part of specimen) are established thanks to it. Using the voucher/accession number attributed in the museum collections as the ID will allow to retrieve information of each linked object within DaRWIN. If you use your own IDs, this link will not be possible...

Values

No values are required. Nevertheless, remember that this tool helps you to import data into DaRWIN, as collection management system. Incomplete information is therefore of limited interest and relevance.

Only use special characters (=, +, /, *, &, #, \$, etc.) in Excel cells **if strictly necessary**, neither as first character nor in subsequent characters. Excel is a data analysis software, therefore it will try to interpret these characters and this may raise errors during export to the XML ABCD formatted file.

Note: Since the use of “.” (dot) in specimen ID is frequent, several tests were made to see if an error could occur. It appears from these tests that the use of a dot surrounded by letters is not a problem. Example: INV.2367. The collection manager can define a default prefix in the collection settings. Only the numeric part should therefore be specified as specimen ID, the prefix being automatically added during the import in the collection.

If you don't have information for a cell, leave it blank. This should limit the presence of uninformative values in your exported XML ABCD formatted file or possibly the number of errors to correct during the import into DaRWIN.

For some columns, the program expects **specific formats or predefined values** given in the list of supported fields available in the appendix of this document. If this format is not respected, the value cannot be taken into account or replaced by default values and you could end up with errors or unexpected values in your exported XML ABCD formatted file.

Some good practices when you fill in the template:

- **People name**
 - o **The more complete the name is, the better.** You should give the first name, the last name and the title.
 - o For the same person, **always use the same spelling.**
 - o **Avoid irrelevant values** as "anonym." or "NA".
 - o **You can reference more than one person:** their names should be separated by a semicolon.
- **Sampling code: One code is used for one collecting event.** A collecting event is defined by its code, date (from/to) and geographic coordinates. Based on these elements, one of the GTUs in the template could match a pre-existing one in DaRWIN for these fields. This pre-existing one will be linked to your specimen and edited in order to add additional information that could be specified in the template (in one of the fields included in the "Collecting event" from the list of supported fields, available in the appendix of this document). If you wish to create a new GTU and not use/edit an existing one, you have to use a different combination of code, date and geographic coordinates...
- **Taxonomy:** The version "_taxonFullName" of the template is preferred. By preparing the Taxonomy Catalogue before the import, the number of errors to correct during the import should be reduced, and you will only have to give the taxon name or the lower known taxonomical name in the template (_taxonFullName version). By using "classical" version of the template with explicit taxonomical levels, only exact match in names and hierarchy will make you avoid correcting errors. More details in the "December 2014 Updates" section.
- **Dates should always be complete.** Supported formats are: YYYY, YYYY-MM and YYYY-MM-DD. If you only have incomplete information or if you want to add textual part in your date, use the fields dedicated to comments.

TAXONOMY-sheet (only in the "classical" template)

If you wish to store your taxonomical levels higher than the family name in the TAXONOMY-sheet, you just have to fill in the TAXONOMY-sheet and be careful to have the family names in both SPECIMEN and TAXONOMY sheets.

Use the "Complete Taxonomy" tool in the custom "ABCDshema" menu that will copy information from the TAXONOMY-sheet to the SPECIMEN-sheet, where the phylum, the class and the order are all missing for the specimen, and the family name is recognized.

Checks

You can check if your latitudes/longitudes are well-formatted, the presence of duplicated IDs and the correct structure of the template (name for SPECIMEN-sheet and column titles), by using the **buttons in the groups "Tools" and "Checks" in the custom "ABCDschema" menu.**

Use filter (select the heading row, click on "Filter" in the "Data" menu) to check your values. You can **see whether the expected values or formats were used.**

One template = one xml file = import in one collection in DaRWIN

It is not possible to import only a part of records from one xml ABCD formatted file in one collection and the remaining records from the same file in another collection. Once the xml file is created, each record it contains can be imported in one unique collection.

If some specimen are not yet published or should not be visible for everybody, they should be stored in another template and imported in a private collection. You can always transfer these specimens to another collection once they are published. You can for example create a collection and add a public sub collection and a private one, that can be grouped later.

Number of “exportable” rows

The template should not contain more than 3000 rows. If you wish to import more records than 3000 in one collection DaRWIn, you should split this dataset into different templates of 3000 rows.

TEMPLATE STRUCTURE

The template for encoding consists of:

- the worksheet SPECIMEN containing data with regard to the specimen (collecting, identification, etc.) where the “specimenID” column is required (even if the cells are left blank)
- only in the “classical” template: the worksheet TAXONOMY where the phylum, the class and the order of the specimen can be stored if you wish. Thanks to the storage of the family names in both the SPECIMEN and the TAXONOMY sheets, the phylum, the class and the order can then be copied, just before exporting to the XML ABCD formatted file (purpose: avoid recopying these values in each row).

This template was designed to minimize the requirements when encoding, which supposes that your data is clean. This allows for a large range of data that can be encoded, but also implies that the values encoded in the Excel cells will be exported as such. In some cases, for example in fields containing date, character strings or alphanumerical data could raise errors during the import.

During the export, the macro will look for which information is stored in the template thanks to the title row (located in the second row of the SPECIMEN-sheet). Only data stored in columns where the title was recognized will be exported to the XML ABCD formatted file. Errors will occur if the titles are not well spelled and/or not in the second row of the worksheets. The order of columns doesn't matter. You can add as many columns as you wish, keeping in mind that they will not be recognized and thus the information they contain will not be exported to the XML ABCD formatted file.

For more information about the format restrictions and correct title spelling for each field, see the list of supported fields available in the appendix of this document.

A tool was added to verify the correct mapping of columns in the template before export. It will tell you which columns are not recognized and if the SPECIMEN-sheet is well found. If the required IDs column ('specimenID') is missing the export will be stopped and an error will be raised.

A warning message will also pop up if a column is not recognized, telling you which headers are concerned. You can decide to go on with the export, by clicking “Yes”: the program export your data without taking unrecognized columns into account. You can abort the export, by clicking on “Cancel” in case of misspelling...

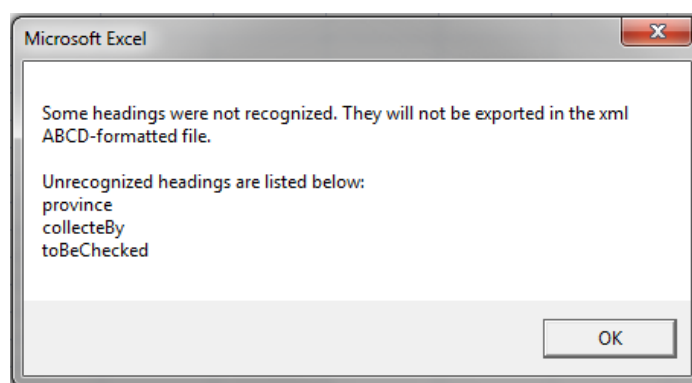


Figure 5 - Result of columns mapping

In this example (Figure 5):

- toBeChecked could be additional information, for internal purpose
- collecteBy is obviously misspelled and should be corrected to collectedBy
- province is available for the export, but the correct name for this field is state_province

Except columns for IDs, you can decide to keep only the relevant columns for the data you want to store in the template. The presence or absence of columns is completely customizable. For example, if you never mention the ocean or the sea where you collected your specimens, these columns do not have to be present in your template.

ADDITIONAL INFORMATION TO FILL THE TEMPLATE

Some requirements or limitations for filling the template are necessary in order to concur with the ABCD schema or with the DaRWIN structure. They are listed in the following paragraphs.

1. Introduction

You don't have to complete each cell. No values are required. Nevertheless, remember that you use this template to import your data into DaRWIN, as collection management system. Incomplete information is therefore of limited interest and relevance, for you and for any other scientist.

If you don't have any information for a cell, leave it blank. For example, when you don't know each level for taxonomy, do not add a dash or a question mark, leave the cell empty. The macro doesn't make the difference between real values and a dash, a question mark, "NA", etc. Consequently, it will export the exact character string that it reads in the cells and you will end up with this uninformative values in your exported XML ABCD formatted or possibly with more errors to correct during the import into DaRWIN.

Only use special characters (=, +, /, *, &, #, \$, etc.) in Excel cells **if strictly necessary**, neither as first character nor in subsequent characters. Indeed, Excel is a data analysis software. It will therefore try to interpret these characters and this may raise errors during export to the XML ABCD formatted file.

Nevertheless, as the use of a dot (".") in specimen ID is frequent, several tests were made to see if an error could occur. It appears from these tests that the use of a dot surrounded by letters is not a problem. Example: INV.2367. Remember that it will be stored as such in the "code" field of the Codes Widget, and not split into a prefix and a code. If you wish to add a prefix for one collection, the curator of the collection should edit the collection to explicitly give a default prefix to the code.

For some columns, the program expects specific formats or predefined values given in the list of supported fields available in the appendix of this document. If this format is not respected, the value cannot be taken into account or can be replaced by default values and you could end up with errors or unexpected values in your exported XML ABCD formatted file.

Using the fields for other kinds of information than what is expected will result in errors being raised or may lead to irrelevant information in the database.

2. ID number

a. Column names for IDs

For each row, an ID should be present. This ID should be unique and correspond to the voucher ID/accession ID attributed to the object in the museum collections. Nevertheless, for specimen that you don't own, you could possibly not know the ID. You can therefore use a custom ID, to allow cross-referencing between the information about the specimen and the samples.

Following the definition of the ABCD concept "UnitID", *"The unit ID should provide the key by which a specimen or specimen component can be identified. Preferably, the unit ID should be stable in the database, so that it also can be used to find the same record again (e.g. for data exchange purposes)."*¹

The identification number of an object is consequently used to distinguish the object from other objects in the collection or department. It should be unique within the collection. This uniqueness also allows to build relationships between objects.

¹ <http://wiki.tdwg.org/twiki/bin/view/ABCD/AbcdConcept0140>

Within an institution, these unique identification numbers can be completed by an additional identifier, as the acronym of the institution and the collection or department to which it belongs. In this way, uniqueness is extended within the institution and even beyond the institution.

One **suggested** format to construct your identification numbers is the following, **but you can use the format that suits you the best**: [YYYY]_[CollectionOrDatasetCode]_[SubGroup]_[Iterative_nb]. The year should consist of 4 digits. The collection or dataset code may include an acronym representing the expedition and/or the institutional registration number. The subgroup may be the name or an acronym of the order/family concerned. The iterative number is a unique number in the collection/the department or the subgroup. Example: 2013_RBINS23134_AVES_01034.

Remember that an unique identifier (ID) will be attributed to each encoded specimen in the DaRWIN database at the moment the specimen is created. This ID is guaranteed to be unique and stable among the whole DaRWIN database. This database ID is not to be confused with the specimen ID that is used within departments and that is not guaranteed to be unique, even if more convenient to use for scientists and curators. The specimen ID is imported as a code, with the category “main”, in the DaRWIN database. If the collection curator defined a default prefix and/or suffix for the collection, you just need to write the numeric part of the code in the template, and the prefix and/or suffix will be automatically added during import. Remember that the whole content of the specimen ID will be present in the field “Code” of the “Codes” widget in DaRWIN. Example: INV.2367 in the specimenID field will be stored in the “code” field in the “Codes” widget and not split into a prefix and a code. To have such a subdivision, the prefix “INV” has to be defined as default for the collection by the curator, and sole 2367 should be present in the specimenID field.

b. Duplicated IDs

Duplicates IDs are not allowed in the “specimenID” column. You can check the presence of duplicates for these by using the corresponding tool available in the ABCDschema menu. The reason behind that is if an association is set on one unit thanks to its ID, duplicated IDs could lead to irrelevant associations in DaRWIN (association with several records that have the same ID).

3. Taxonomy

More information is available in the “December 2014 Updates” section of this manual.

By preparing the taxonomy corresponding to your specimens before the specimens import, you should limit the number of errors occurring during the import. The basic principle is the “exact match”, that is:

- for the “_taxonFullName” version of the template, only exact match between “taxonFullName” column and DaRWIN taxonomical entry
- for the “classical” version of the template, exact same name and hierarchy between the template and DaRWIN.

We therefore strongly advise to use the Method 1 with the “_taxonFullName” version of the template.

If you prefer to use the “classical” version of the template without preparing and/or checking the taxonomy before specimens import, you can fill as much information as you wish in the fields for taxonomy. The more the tree is filled in, the easier it should be to check and create taxonomy in DaRWIN during the import. The matching with the DaRWIN database will nevertheless not occur frequently and you will be asked to select/create the correct entry. The “taxonFullName” column in this template is dedicated to the comparison of the name of the taxon from genus level or below for your specimen with entries DaRWIN. It should be filled in to ensure a correct matching of the existing taxon. If this field is left empty, it will be automatically created based on the genus, subgenus, species, subspecies and variety_form and author_year fields without warranty that it will correspond to your expectations. Warning: the completion of the ‘taxonFullName’ field do not replace the genus, subgenus, species, subspecies and variety fields. They should all be completed.

It is currently not possible with the import tool to add a vernacular name for a taxon. This name should be added separately within DaRWIN in the taxonomical unit.

The taxonomy in DaRWIN follows a strictly defined structured and does not allow a follow-up of taxonomy. There is a high risk of errors when using expressions like « Various amphipods » or « sp1 » and thus the result of import cannot be guaranteed. In the case of an uncertain/incomplete determination, the good DaRWIN practice is to fill in only the taxonomy levels you are sure of

and to leave the rest blank. For example, if the genus/species for a specimen is unknown/undescribed, you would have to fill in the taxonomical tree up to the family level, and leave genus and species blank. Expressions like « [family name] gen. n. det. » are not to be used in the database. You can add a comment to precise the determination status. Alternatively, it is possible to add a suggestion or report a problem in the 'suggestions/report problem' widget, directly in the DaRWIn interface. By doing so, the specimen is flagged, can be easily monitored, and when possible, updated or corrected.

4. TAXONOMY-sheet (only in “classical” template)

The TAXONOMY-sheet is structured in 4 columns: phylum, classis, ordo and familia. If one of these columns is missing a warning message will appear. A column “familia” is present in the SPECIMEN-sheet as well. If this column is missing, a warning message will also appear. These two ‘familia’ columns allow the program to link the higher taxonomical levels listed in the TAXONOMY-sheet to the specimens in the SPECIMEN-sheet. Be careful that the spelling must be exactly the same in the two sheets. If the columns dedicated to store the phylum, the class and the order are not present in the SPECIMEN-sheet, they will be automatically created and added after the last column of the worksheet.

You are not obliged to fill in the TAXONOMY-sheet. This tool is provided to allow you to write only once the taxonomical tree for each family name instead of having to copy it in every row.

When you wish to copy information from the TAXONOMY-sheet to the SPECIMEN-sheet, click on the “Complete Taxonomy” tool in the custom “ABCDshema” menu that will copy taxonomical information in the SPECIMEN-sheet where the phylum, the class and the order are all missing for the specimen, and the family name is recognized.

5. Dates

For technical reasons, each date in the template is divided into 3 columns: one for the year, one for the month and one for the day.

In the ABCD schema, dates follow the ISO/ANSI 8601 standard structured format. The following formats are available and recognized in the template:

- YYYY: when only the year is known
- YYYY-MM: when only the month and the year are known
- YYYY-MM-DD: when the exact date is known

If this format is not valid (i.e., you did not enter a correct numeric value in the date fields), the macro will try to structure the information and store it as a comment, without any warranty of success. If you only have imprecise information (e.g., “before 2012”), use preferably the columns dedicated to comments.

6. People name

The name should be as complete as possible. You should mention the first name, the last name and the title. Indeed, during the import into DaRWIn, you could be asked to create a new entry for this person in the People Catalogue. You need full information to create a complete entry in this catalogue.

For the same person always use the same spelling. For example, “C. Darwin” or “Darwin, C.” is not considered as the same values by the import tool in DaRWIn. If you don’t know the collector name, the identifier name, etc., leave the cell empty and avoid irrelevant values as “anonym.” or “NA”. The name of the mission/expedition should not be referenced in the columns dedicated to people.

You can reference more than one person in cells designated for such purpose. Their names should be separated by a semicolon. Otherwise, the import tool in DaRWIn will not recognize the presence of more than one name.

7. Sampling code and sampling locations

The field “samplingCode” is used as identifier for the sampling location, similarly to the identifier of a specimen. One code is used for one collecting event. A collecting event is defined by its code, the date (from/to) and the geographic coordinates.

Based on these elements, one of the GTUs in the template could match a pre-existing one in DaRWIN for these fields. This pre-existing one will be linked to your specimen and edited in order to add additional information that could be specified in the template (in one of the fields included in the “Collecting event” from the list of supported fields, available in the appendix of this document). Indeed, all the additional fields included in the “Collecting event” are considered as complementary information. If you wish to create a new GTU and not use/edit an existing one, you have to use a different combination of code, date and geographic coordinates...

If you wish to use an existing GTU for one or more specimens, you should give the same code, date and geographic coordinates. If a matching occur during the import, this GTU will be linked to the specimen and all additional tags, properties or comments will be added to this GTU (through an edition) so that a new GTU will not be created.

If you wish to create a new GTU, you have to use a different combination for code, date and geographic coordinates. For example, two collecting events taking place at the exact same place with the same environmental conditions except for the depth that is different should have two different codes.

8. Container storage

The fields (sub)container, (sub)containerType and (sub)containerStorage correspond to the so-named fields in the Container widget in DaRWIN. While (sub)container is a text field where you can give the name and/or number of the (sub)container, containerType and containerStorage are user-defined lists. Be careful to use same expressions in the template than what is already present in DaRWIN in order to prevent the creation of duplicated type or medium of storage in the drop-down list in the widget.

9. Pre-export checks

You can check the latitude/longitude values and see if they are correctly encoded. They are automatically converted into a decimal format (if not encoded as such) for the ABCD export. If this conversion failed, you will receive a message telling you which values are erroneous. If errors are found, they will be listed in the “CheckLatLong” sheet. You can access this tool “Check Latitude/Longitude” in the group “Tools” of the custom “ABCDschema” menu.

You can also check the presence of correctly named worksheets, the mapping of column names and the presence of duplicated IDs before exporting the data, to allow a correction. These checks are available by clicking the buttons “1. Columns mapping” and “2. Duplicated IDs” in the group “Checks” of the custom ribbon “ABCDschema”. A window will appear, listing the potential problems. It is highly recommended to run these checks before trying the export.

Finally, you can use the filter (select the heading row, click on “Filter” in the “Data” menu) to check your values. This way, you can see if you use only the expected values or formats in the different columns. For example, you can check if there are only numerical values in other columns for dates, if the same names have the same spelling or if fields where only specific values are supported do not contain erroneous values, etc.

In the example below (Figure 6), instead of a numeric value for the year of identification event, the values “before 1950” and “<1939” were written. Such values could not be present in the exported XML ABCD formatted file, or they could raise errors. Thanks to the filter, you can display only rows with these values, and correct them. In this case, this information could be stored in a comment for example.

TECHNICAL INFORMATION – ABCDSchema TAB

1. Commands

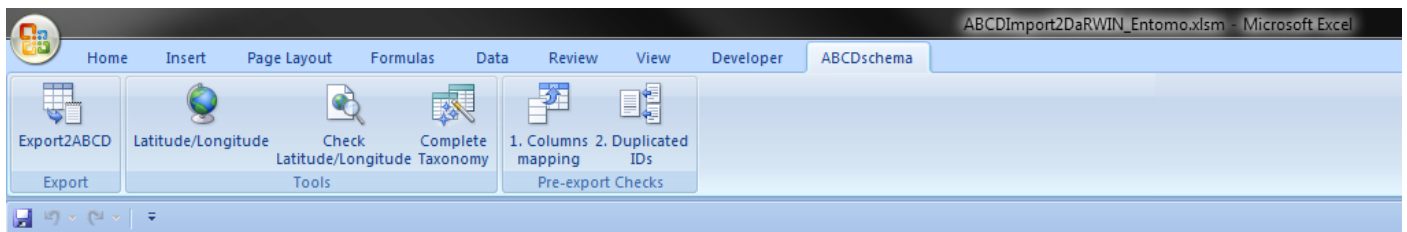


Figure 7 - ABCDSchema Menu

a. Export group

- *Export2ABCD* : creates XML file that matches the ABCD schema with the data contained in the Excel file.

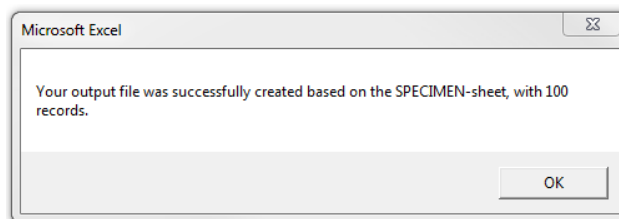


Figure 8 - Result of the export to XML ABCD formatted file

b. Tools group

- *Latitude/Longitude* : helps to insert latitude/longitude with a correct format.

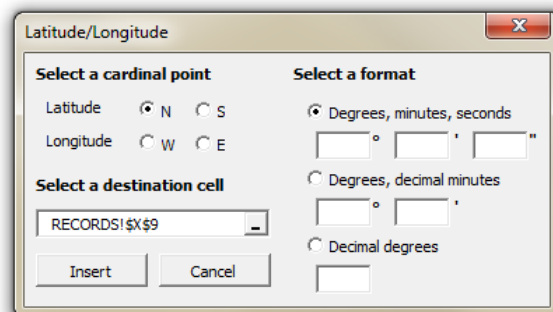


Figure 9 - Latitude/Longitude Tool

- *Check Latitude/Longitude*: helps to check if your latitude/longitude were correctly encoded. It creates a worksheet named "CheckLatLong" with a listing of erroneous values (and their mapping in the worksheet "SPECIMEN"). If no errors are found, a message tells you that everything looks OK.

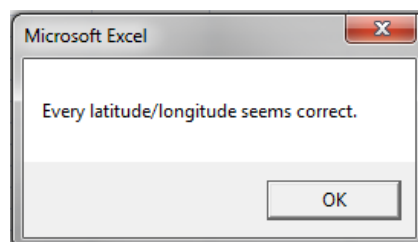


Figure 10 - No errors detected in Latitude/Longitude.

- *Complete Taxonomy*: allows you to copy information about higher taxonomical levels from the TAXONOMY-sheet to the SPECIMEN-sheet. This will work only if:
 - ✓ the family name referenced in the SPECIMEN-sheet has a corresponding value (with the same spelling) in the TAXONOMY-sheet
 - ✓ the cells for higher taxonomical levels (phylum, class and order) are blank in the SPECIMEN-sheet.

c. Checks group

- 1. *Columns mapping* : checks if each column title is recognized and if the SPECIMEN-sheet and the 'ID' column are found.
- 2. *Duplicated IDs* : checks if no duplicated IDs are present.

2. Output

a. Name and extension

[USERDEFINED_NAME].XML

The export will create a file with extension .xml. You can choose the name and the folder where you wish to store this XML ABCD formatted file.

b. Structure

Globally, units are encoded step by step, looping within the SPECIMEN-sheet. One unit is created for each row.

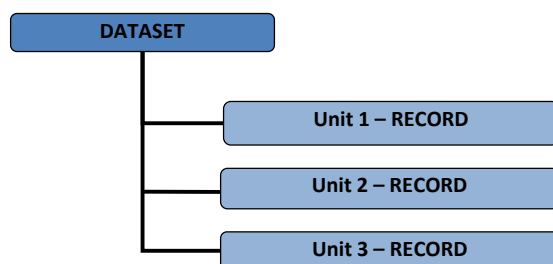


Figure 11 - Export XML file structure

GLOSSARY

XML and ABCD XSD schema

XML stands for eXtensible Markup Language and is a markup language much like HTML.

XML was created to structure, store, and transport information. Documents are therefore encoded in a format that is both human-readable and machine-readable. It is mainly used as intermediate format between two computers or softwares.

Information is structured through tags. These tags surround information by creating one element. Different elements can be fitted together, to create a hierarchical structure. A tag begins with "<" and ends with ">". Each element has a start-tag `<tag>` and end-tag `</tag>`. Empty tags takes the form `<tag />`.

XML Schema can be used as template for structuring information in your XML file. It rigorously defines the structure of your document. ABCD (Access to Biological Collection Data) is a predefined format to store biodiversity collections, developed by TDWG (Taxonomic Database Working Group). Several extension were developed:

- Extension for Geosciences (EFG)
- Extension for DNA data (ABCDDNA)
- Extension for herbarium collections (HISPID)

Visual Basic For Applications and macros

Excel has a language called VBA (Visual Basic for Applications). This language enables to program excel to automate several tasks. A macro is nothing but a set of instructions you give Excel in the VBA language.

The code for exporting your data filled in the template into an XML ABCD formatted file was prepared in a macro using VBA.

APPENDIX

1. General list of supported fields with expected format, description and example(s)

In the SPECIMEN-sheet, the following fields are supported corresponding to the information that can be retrieved in DaRWIn after import. Warning : classification, phylum, classis, ordo, superfamilia, familia, subfamilia, genus, subgenus, species, author_year, subspecies, variety_form are not available in the “_taxonFullName” version of the template.

Field	Format	Description and comments	Example
specimenID		Unique identifier of specimen if available. Example of format: [YYYY]_[CollectionOrDatasetCode]_[SubGroup]_[Iterative_nb]. The year should consist of 4 digits. The collection or dataset code may include an acronym representing the expedition and/or the institutional registration number. The subgroup may be the name or an acronym of the order/family concerned. The iterative number is a unique number in the collection or the subgroup.	RBINS_ENT_Term2678
code		Additional identifier, for internal purpose only (classification of types, etc.)	Termites001
datasetName		Name or code for the project, expedition, etc. as complementary information for the collection name, choosen in DaRWIn at the moment of the import	LubumbashiExp2013
accessionNumber		IG Number in RBINS: institutional number given to each new group of items acquired by the institution and recorded in the collection registers	25442
phylum		Phylum	Arthropoda
classis		Class	Insecta
ordo		Order	Isoptera
familia		Family	Termitidae
genus		Genus	Macrotermes
species		Species	falciger
author_year		Author, year or (Author, year) for the species	(Meyer, 1999)
subspecies		Subspecies	goliath
taxonFullName		This field is dedicated to the taxon full name, as you wish to see it in the Taxonomy Catalogue in DaRWIn. If you do not fill in this field, the taxon full name will be recomposed based on the data in the genus, subgenus, species, subspecies fields but this could not match what you expect.	Example 1 : Acmaeodera iridea Kerremans, 1902; Example 2 : Abeillia abeillei abeillei (Lesson & Delattre, 1839)
identifiedBy		Identifier name (title)	Samyn Yves (Dr.)
identificationNotes		Additional information/remark about the taxonomy	Uncertain determination, atypical specimen
samplingCode		A code that identifies all material or records resulting from a collecting or observation event	TOL201304
continent		Continent (administrative name)	Africa
country		Country (administrative name)	Madagascar
state_province		State or province, as a subdivision of a country (administrative name)	Southwestern Coast
region_district		Region or district (administrative name)	Atsimo-Andrefana
municipality		Locality (administrative name) or urban administrative division	Toliara
exactSite		Site name, alternative name, lieu-dit, how many kilometers and compass direction from the nearest major specific map location (e.g. town, mountain peak, lake, specific park or refuge, etc.), road network. All distances should be presented in metric units.	Reef Toliara
latitude	Decimal degrees preferred (DD°MM'SS" or DD°MM.MM' also accepted)	Latitude (WGS84) - use N, S, E or W for cardinal points	10°58'53"S
longitude	Decimal degrees preferred (DD°MM'SS" or DD°MM.MM' also accepted)	Longitude (WGS84) - use N, S, E or W for cardinal points	26°44'12"E
ecology		Biotope or any ecological information related to the collecting site	termite mound
samplingMethod		Material and/or method used for the sampling event	net M5-F4
collectedBy		Collector name (title)	Sonet Gontran (M.)
collectionStartDay	Numeric, 2 digits	Day of the unique date or starting date of collecting event	12
collectionStartMonth	Numeric, 2 digits	Month of the unique date or starting date of collecting event	4
collectionStartYear	Numeric, 4 digits	Year of the unique date or starting date of collecting event	2013
collectionEndDay	Numeric, 2 digits	Day of the ending date of collecting event	16
collectionEndMonth	Numeric, 2 digits	Month of the ending date of collecting event	4
collectionEndYear	Numeric, 4 digits	Year of the ending date of collecting event	2013
localityNotes		Additional information/remark about the collecting event	Examples: Collected before 2006; collected in the morning; ebb tide; collected in some traps trampled or transect along a west-east gradient of 10 miles; specimen collected dead
kindOfUnit		Part(s) of organism or class of materials represented (animal, mounted specimen, microscopic preparation, part of organism, etc.)	specimen
statusType		Type (paratype, holotype, etc.)	Paratype
socialStatus		Social status	winged adult
totalNumber	Numeric	Total number of specimens or part of specimens in the record	20
adultCount	Numeric	Number of adults	3
larvaCount	Numeric	Number of larva	3
pupaCount	Numeric	Number of pupa	3

maleCount	Numeric	Number of males	2
femaleCount	Numeric	Number of females	4
sexUnknownCount	Numeric	Number of specimens with unknown sex	5
conservation		Chemicals used or physical process for long term conservation	Ethanol
institutionStorage		Institution storage	Royal Belgian Institute of Natural Sciences
buildingStorage		Building storage	De Vestel
floorStorage		Floor storage	6
roomStorage		Room storage	28
laneStorage		Row storage	2
columnStorage		Column storage	1
shelfStorage		Shelf storage	1b
container		Container name and/or number	126
containerType		Type of container	Example 1: Jar; Example 2: box
containerStorage		Conservation mean	Example 1: dry; Example 2: alcohol
subcontainer		Subcontainer name and/or number	AI
subcontainerType		Type of subcontainer	Example 1: slide; Example 2: minigrip
subcontainerStorage		Subcontainer mean	Example 1: dry; Example 2: alcohol
notes		Additional information/remark about the specimen that doesn't fit elsewhere	Example 1: Not for exhibitions; Example 2: Damaged due to flooding

In the TAXONOMY-sheet, the following columns are supported: phylum, classis, ordo and familia and are all required.

2. Link between old and new field names

Old name	Field
ID_Number	specimenID
Code	code
Collection	datasetName
IG	accessionNumber
Phylum	phylum
Class	classis
Order	ordo
Family	familia
Genus	genus
Species	species
Author_year	author_year
Subspecies	subspecies
	taxonFullName
Identifier_name	identifiedBy
	identificationNotes
	samplingCode
	continent
Country	country
Region	region_district
State	state_province
Locality	municipality
Toponym	exactSite

Old name	Field
Latitude	latitude
Longitude	longitude
Ecology	ecology
Lure	samplingMethod
Collected_by	collectedBy
Collecting_start_D	collectionStartDate
Collecting_start_M	collectionStartMonth
Collecting_start_Y	collectionStartYear
Collecting_end_D	collectionEndDay
Collecting_end_M	collectionEndMonth
Collecting_end_Y	collectionEndYear
Locality_notes	localityNotes
Kind_of_Unit	kindOfUnit
StatusType	statusType
Social_status	socialStatus
Total_number	totalNumber
N_adult	adultCount
N_larva	larvaCount
N_pupa	pupaCount
N_males	maleCount
N_females	femaleCount
N_sex_unknown	sexUnknownCount

Old name	Field
Conservation	conservation
	institutionStorage
Building	buildingStorage
Floor	floorStorage
Room	roomStorage
Row	laneStorage
Colon	columnStorage
Shelf	shelfStorage
Box	-
Tube	-
	container
	containerType
	containerStorage
	subcontainer
	subcontainerType
	subcontainerStorage
Remark	notes

The Subcollection is not supported anymore (for technical reasons, not supported in DaRWIn).