

# USER MANUAL

---

## TEMPLATE ABCDIMPORT2DARWIN\_DNA: ENCODING DNA DATA IN XML-FILE (ABCD SCHEMA STRUCTURE)

Version		Author(s)	Date	Description
1.0	ABCDImport2DaRWIn_DNA_v1_0.xlsm	MAdam	January 2014	Testing version of the template for import
1.1	ABCDImport2DaRWIn_DNA_v1_1.xlsm	MAdam	March 2014	Production version of the template for import
1.2	ABCDImport2DaRWIn_DNA.xlsm = <b>"CLASSICAL" version</b>	MAdam	August 2014	Small adaptations and debugging
1.2 bis	ABCDImport2DaRWIn_DNA_taxonFullName.xlsm = <b>"_taxonFullName" version</b>	MAdam	December 2014	New version with simplified taxonomy

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

Extension for DNA data (ABCDDNA): <http://www.dnabank-network.org/>

## TABLE OF CONTENTS

<b>December 2014 – Updates .....</b>	<b>3</b>
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
<b>September 2014 – Updates .....</b>	<b>5</b>
<b>User information .....</b>	<b>6</b>
<b>Template structure.....</b>	<b>9</b>
<b>Additional information to fill the template .....</b>	<b>10</b>
1. Introduction .....	10
2. Ownership.....	10
3. ID number .....	11
a. Column names for IDs .....	11
b. Associations between specimens, samples and DNA-extracts .....	11
c. Duplicated IDs .....	12
4. Taxonomy .....	12
5. Dates .....	13
a. General format .....	13
b. DNA extraction date .....	13
6. People name .....	13
7. Sampling code and sampling locations .....	13
8. Custom site and specimen properties .....	14
9. Container storage in the SPECIMEN-sheet .....	14
10. Pre-export checks .....	14
<b>Technical information – ABCDschema TAB .....</b>	<b>16</b>
1. Commands .....	16
a. Export group.....	16
b. Tools group.....	16
c. Checks group .....	17
2. Expandable columns in SPECIMEN-sheet .....	17
3. Output.....	17
a. Name and extension.....	17
b. Structure.....	18
<b>Glossary .....</b>	<b>20</b>
XML and ABCD XSD schema .....	20
Visual Basic For Applications and macros .....	20
<b>APPENDIX.....</b>	<b>21</b>
1. General list of supported fields .....	21
2. List of fields by worksheet, with expected format, description and example(s) .....	22
a. SPECIMEN-sheet .....	22
b. SAMPLE-sheet .....	25
c. DNA-sheet .....	25
3. Link between the fieldsheet and the template .....	26
4. Link between the template and the BOLD Systems Submission format .....	29

## 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, 2 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

	<b>Method 1</b> <i>ABCDImport2DaRWIN_General _taxonFullName.xlsm</i>	<b>Method 2</b> <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 in the SPECIMEN-sheet:
  - 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 still available in the SAMPLE-sheet and the DNA-sheet, but not in the SPECIMEN-sheet anymore since there was a conflict with these new fields.

- Increase of the number of site properties to 10 and of the specimen properties to 20 + small debugging
- New fields for sampling location, for extended possibilities
- Improvement of the export of time for the collecting event (character strings should also be exported successfully)
- Add possibility to reference more than one external link or more than one picture/related file (have to be separated by “;”)

Note: For RBINS users, if your pictures/related files are stored in datastore, be careful to reference url as “smb://datastore/darwintmp/...”

- 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”).

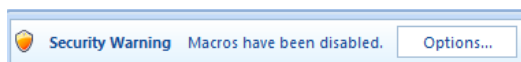


Figure 1 - Security Warning for macros

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.

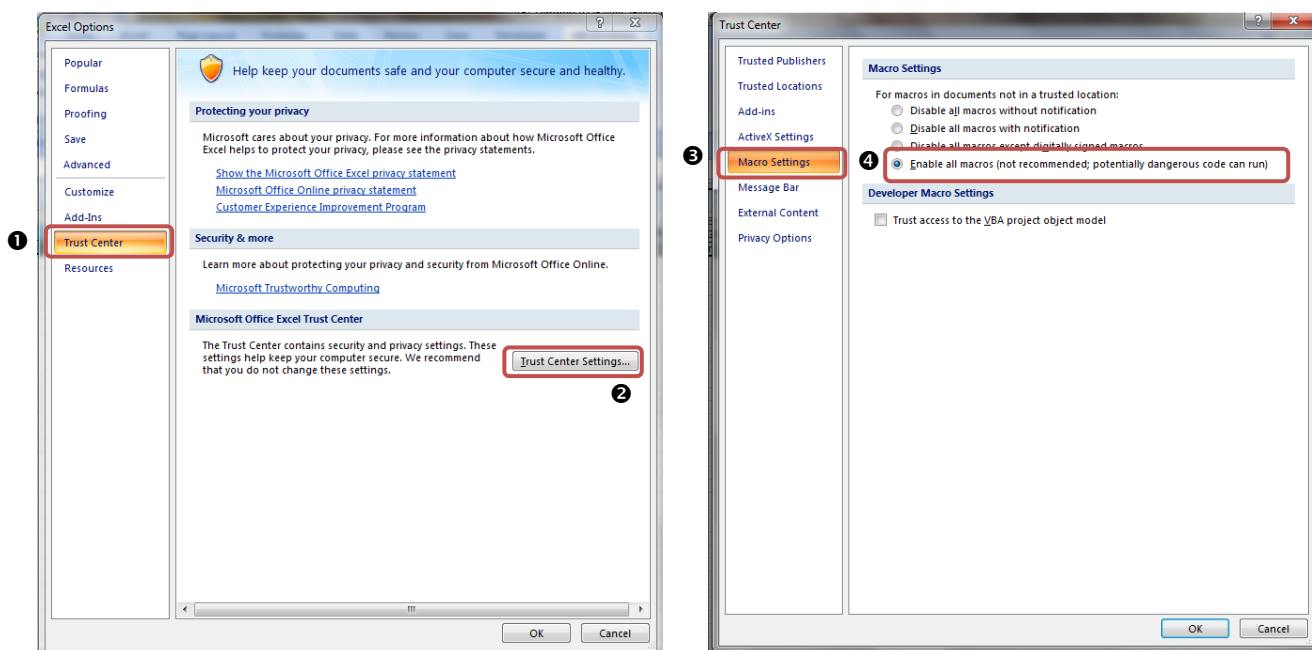


Figure 2 - Excel options: enable macros

### 2. Fill in all worksheets (SPECIMEN, SAMPLE, DNA)

Information about the specimen is stored in SPECIMEN, about the sample in SAMPLE and about the DNA extraction in DNA.

### 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 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” in “ABCDschema” menu 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.

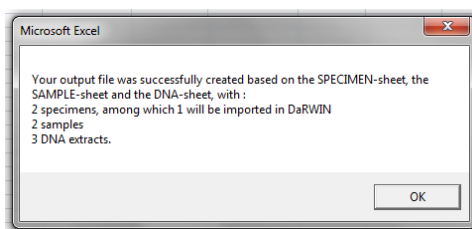


Figure 3 - Summary of your export

## QUICK PRATICAL GUIDE

### Template structure

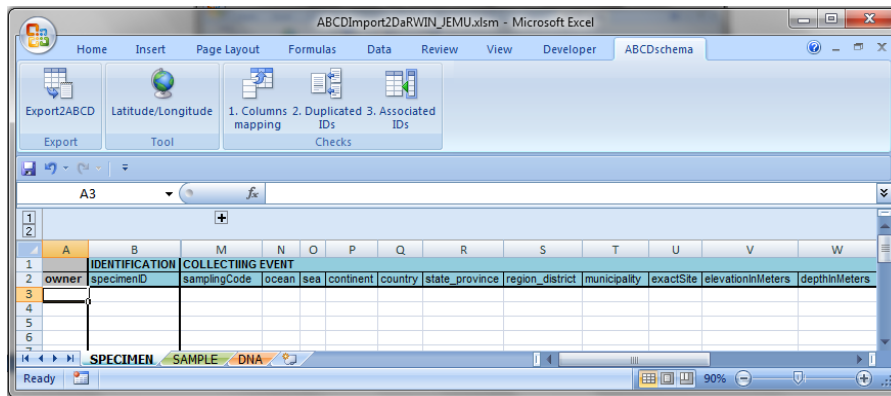


Figure 4 - Template structure

Verify that the following elements are present:

- **3 worksheets, named 'SPECIMEN', 'SAMPLE' and 'DNA'**
- **A title for each column, 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.
- **A column for IDs in each worksheet**, named 'specimenID', 'sampleID' and 'dnaID' respectively  
An ID is not required, but remember that the links between specimens, samples and dna-extracts 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**
  - **The more complete the name is, the better.** You should give the first name, the last name and the title.
  - For the same person, **always use the same spelling.**
  - **Avoid irrelevant values** as “anonym.” or “NA”.
  - **You can reference more than one person** (except for DNA extraction responsible, where only one person can be referenced): 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.

### Checks

You can check the presence of duplicated IDs in each sheet, the recognition of association between units, the correct structure of the template (names for worksheets 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 3 sheets :

- SPECIMEN : data with regard to the specimen (collecting, identification, etc.), where the “specimenID” column is required (even if the cells are left blank)
- SAMPLE : data with regard to the sample (type, protocol, storage), where the “sampleID” column is required (even if the cells are left blank)
- DNA : data with regard to DNA extraction (DNA itself and extraction process), where the “dnaID” column is required (even if the cells are left blank)

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 worksheet) of each worksheet. Only data stored in columns where the title was recognised 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 recognised 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 recognised and if required sheets or columns are missing. If a sheet or required field (IDs columns) is missing the export will be stopped and an error will be raised when running Export2ABCD.

A warning message will also pop up If a column is not recognized, telling you for each worksheet 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...

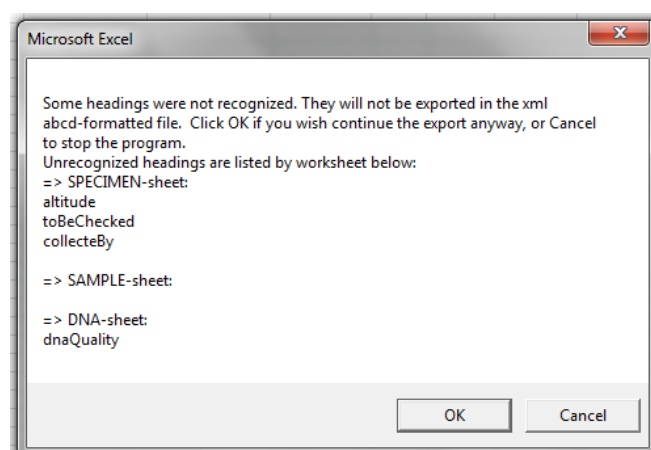


Figure 5 - Result of columns mapping

In this example (Figure 5):

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

- dnaQuality is not in the listing of supported values. Either there is an error and the correct name is “dnaAbsorbance260\_280” for example, or it is additional information that you added in the template, but it will not be exported.

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. Ownership

For each row in the SPECIMEN-sheet, you should mark the cell with “x” in the “owner” column if you are owner of the specimen. This column defines the creation of a record for the specimen in DaRWIn or not, and it will also define what kind of information will be present in the template (see appendix 1 for more details).

If you are owner of the specimen and the owner-cell is checked, all information for this row will be exported, and during the import into DaRWIn, the specimen will be created and affiliated to the same collection as well as the samples and the DNA-extracts.

If you are not owner of the specimen and if the owner-cell is not marked with “x”, only information in dark blue will be exported and stored in the samples and DNA-extracts associated (that is, no specimen will be created in DaRWIn during the import).

If the column “owner” is missing, all specimens are considered as NOT owned by default.

### 3. ID number

#### a. Column names for IDs

For each row in the 3 worksheets, 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)."*<sup>1</sup>

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.

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, the sample ID and the DNA ID are imported as codes, 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, sample ID and DNA 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. Associations between specimens, samples and DNA-extracts

The specimen ID is the key that links the sample to a specimen, and the sample ID is the key that in turn links the DNA extract to the sample ID. It's a hierarchical structure. It is therefore important to add an ID for each object, and take care to fill it in correctly in the different worksheets within the template.

For this purpose, you have to use the column names:

- "specimenID" in the SPECIMEN-sheet,
- "sampleID" and "associatedSpecimenID" in the SAMPLE-sheet, where "associatedSpecimenID" is the ID for the associated specimen written in the "specimenID" column of the SPECIMEN-sheet
- "dnaID" and "associatedSampleID" in the DNA-sheet, where "associatedSampleID" is the ID for the associated sample written in the "sampleID" column of the SAMPLE-sheet

The relation will be written in the imported specimen/sample/DNA extract, but a clickable link to access the view of the associated units will only be created if the specimen AND its associated samples and DNA extracts are imported through the

---

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

same xml file. Otherwise, the relation will be written in the unit view (specimen, sample or DNA extract), and you will have to create the relation manually to make it a clickable link to the associated unit. This functionality should be improved in the future.

A tool was added in the “ABCDschema” menu to check if each association is recognized. A specimen ID present in the SAMPLE-sheet and not in the SPECIMEN-sheet, or a sample ID in the DNA-sheet and not in the SAMPLE-sheet will be highlighted in red.

### c. Duplicated IDs

Duplicates ID’s are not allowed in the “specimenID” column, the “sampleID” column and the “dnaID” 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 that the association is not guaranteed if duplicates exist in IDs of each sheet since the program will scan the values and stop at the first match. Duplicated ID’s could also lead to irrelevant associations in DaRWIn if several records have the same ID.

## 4. 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.

A column “Classification” is present in the SPECIMEN-sheet for the “classical” version of the template. This is necessary for the structuring of the XML file with the correct branch of the ABCD schema. The values allowed are Zoological and Botanical. By default, if this cell is empty or the value misspelled when running “Export2ABCD”, the default value is set to “Zoological”.

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 the specimen are 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.

## 5. Dates

### a. General format

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

For the collecting event, the format YYYY-MM-DD HH:MM is also available, if you know the exact date and time.

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.

### b. DNA extraction date

The extraction date (3 fields : extractionDay, extractionMonth and extractionYear) will only be present in the export file if the 3 fields are completed. Indeed, the ABCDDNA format requires a complete date for validation (YYYY-MM-DD format).

## 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 (except for DNA extraction responsible, where only one person can be referenced) 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. Custom site and specimen properties

You can encode custom site and specimen properties. In other words, in the “siteProperty\_[nb]” or “specimenProperty\_[nb]” column, you can specify which is the measured or described parameter for the site (e.g., temperature) or the specimen (e.g., weight) and precise its value (e.g., 28°C or 230 g) in the corresponding “sitePropertyValue\_[nb]” or “specimenPropertyValue\_[nb]” column. There are five custom properties for the collection site and ten for the specimen.

## 9. Container storage in the SPECIMEN-sheet

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.

## 10. 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.

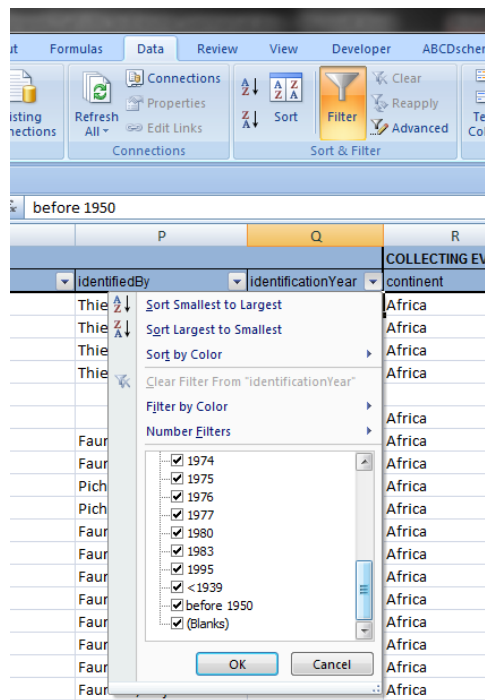


Figure 6 - Filter data in Excel

## TECHNICAL INFORMATION – ABCDSchema TAB

### 1. Commands

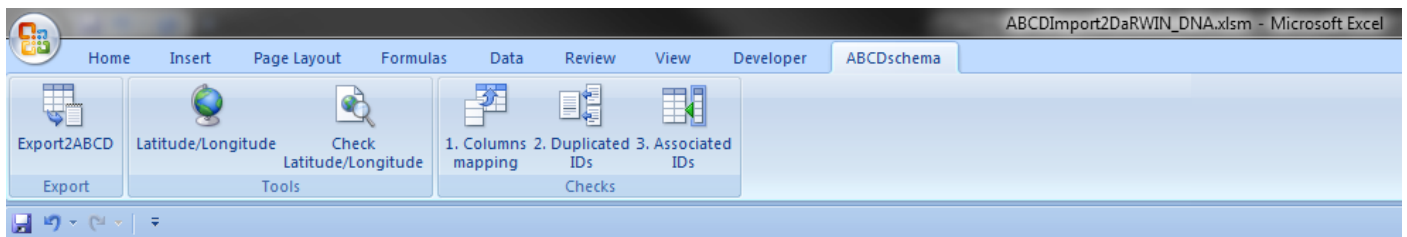


Figure 7 - ABCDSchema Menu

#### a. Export group

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

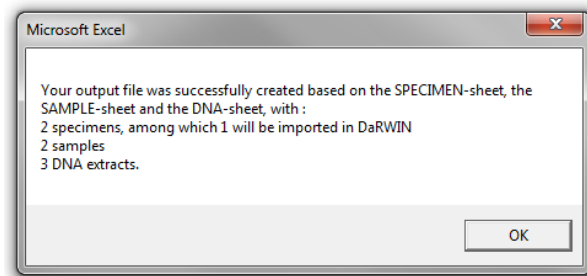


Figure 8 - Result of the export to XML ABCD formatted file

#### b. Tools group

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

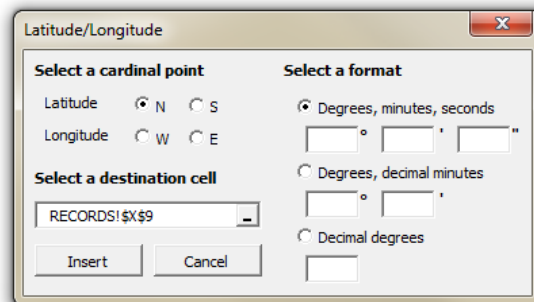


Figure 9 - Latitude/Longitude Tool

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

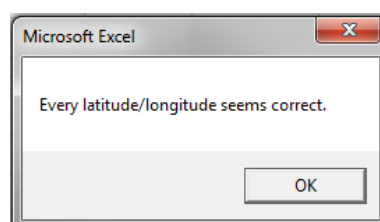


Figure 10 - No errors detected in Latitude/Longitude.



### c. Checks group

- 1. *Columns mapping* : checks if each column title is recognized and if every required sheet/field is present.
- 2. *Duplicated IDs* : checks if no duplicated IDs are present:
  - o in the specimenID column in the SPECIMEN-sheet,
  - o in the sampleID in the SAMPLE-sheet
  - o and in the dnaID in the DNA-sheet.
- 3. *Associated IDs* : checks if every “associatedSpecimenID” in the SAMPLE-sheet is found in the “specimenID” column in the SPECIMEN-sheet and if every “associatedSampleID” in the DNA-sheet is found in the “sampleID” column in the DNA-sheet.

## 2. Expandable columns in SPECIMEN-sheet

The supported fields available for each specimen will be different if you own the specimen or not. If you own the specimen and check the cell in the “owner” column, all fields are supported. If you do not own the specimen, only a limited range of fields are supported (fields highlighted in dark turquoise, see appendix 1 for more details). You can choose to display only the limited supported fields or all of them by clicking respectively on the small “1” or on the small “2” above the grid and the header row (see figure below).

When you click on 1, only limited fields are displayed:

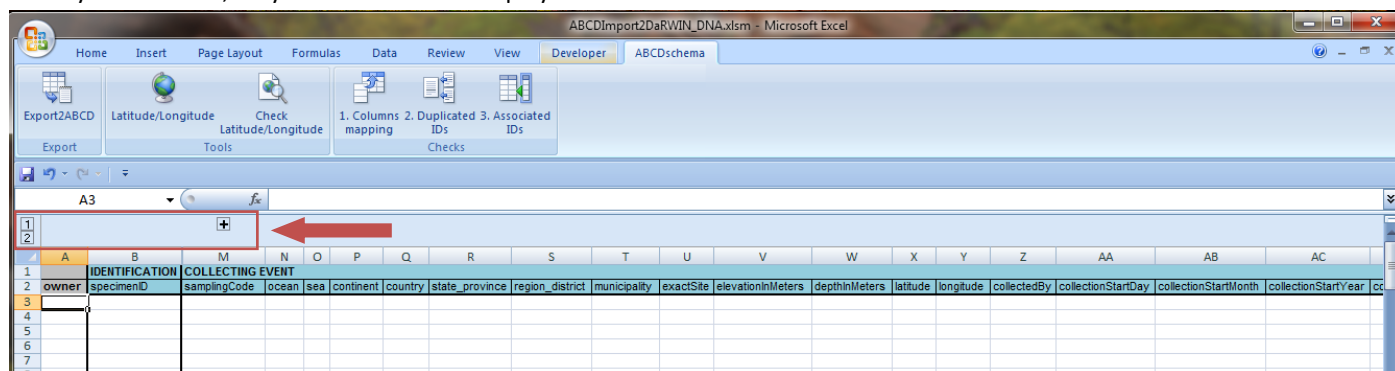


Figure 11 - Limited fields display

When you click on 2, all supported fields are displayed:

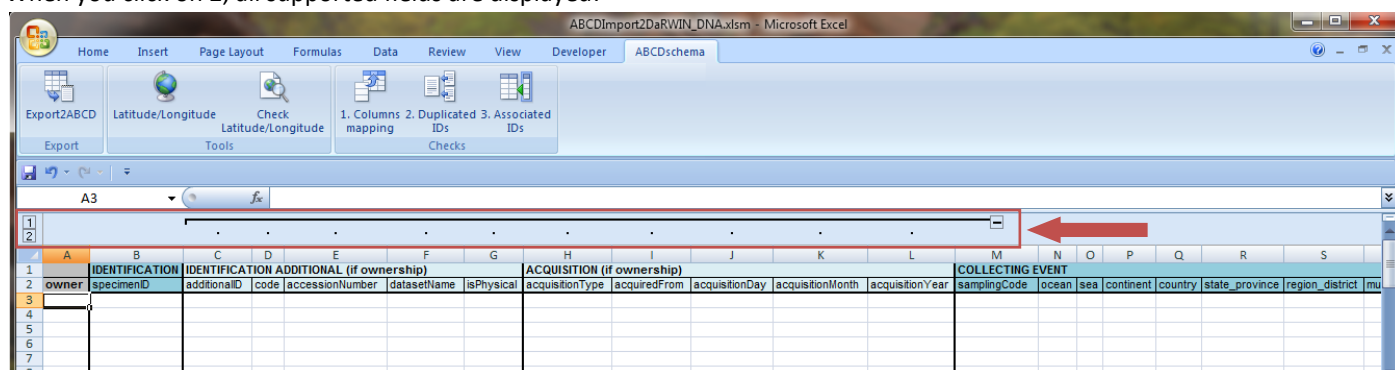


Figure 12 - Full display

## 3. 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

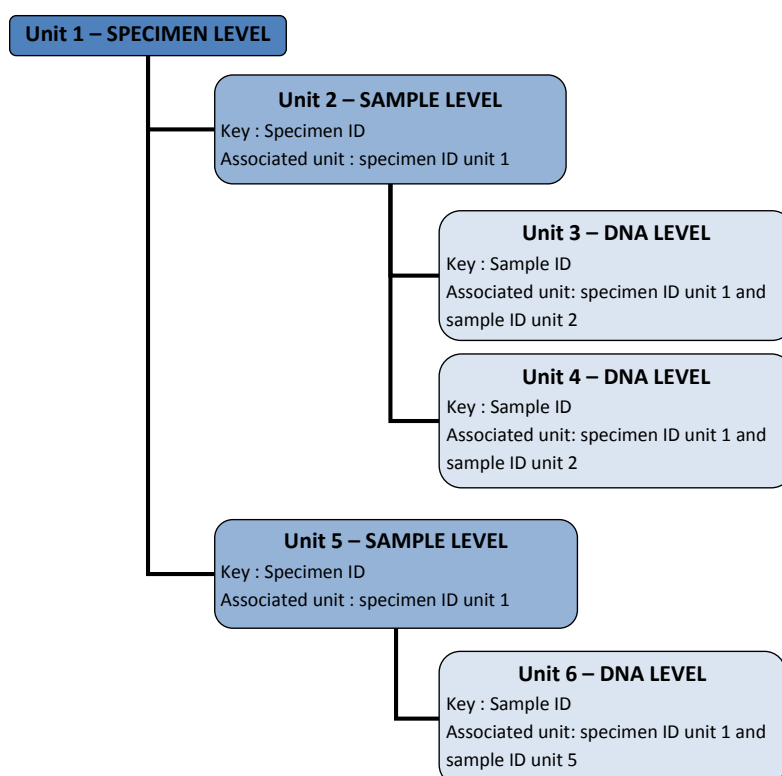
Globally, units are encoded step by step:

- 1) The information about a specimen in the SPECIMEN-sheet is read or exported. The specimen ID is stored by the macro.
- 2) The macro uses this stored specimen ID to find associated samples in the SAMPLE-sheet.
- 3) If an associated sample is found, its information from the SAMPLE-sheet is read or exported. The sample ID is stored by the macro.
- 4) The macro uses this stored sample ID to find associated DNA-extracts in the DNA-sheet.
- 5) If an associated DNA-extract is found, its information from the DNA-sheet is read or exported.
- 6) The macro continues to search for other DNA-extracts associated to the stored sample ID, and reads and exports their information from the DNA-sheet, until no more associated DNA-extract for the sample is found in the DNA-sheet.
- 7) The macro then continues to search for other samples associated to the stored specimen ID, and reads and exports their information from the SAMPLE-sheet, until no more associated sample for the specimen is found in the SAMPLE-sheet.
- 8) For each associated sample, the sample ID is stored and the macro tries to find associated DNA-extracts in the DNA-sheet and reads and exports their related information.
- 9) Once no more associated sample is found for the stored specimen ID, the macro goes to the next line in the SPECIMEN-sheet and the process can restart.

The most complete structure consists of 3 hierarchical levels: specimen, sample and DNA-extract. Each sample is linked to a specimen, and each DNA-extract is linked to a sample (see units 1 to 6 in the figure below).

When you do not own the specimen, only sample and DNA-extracts units will be created, copying information from the specimen in the sample(s) and DNA-extract(s) related to this specimen (example: unit 7 in the figure below).

If association between units are not recognized (i.e., there is a misspelling in the ID), units will be created but they will not be linked to each other (example: unit 10 to 14 in the figure below). In each level, objects can be encoded without associated unit so that no data will be lost but the hierarchy in 3 levels will consequently be incomplete.



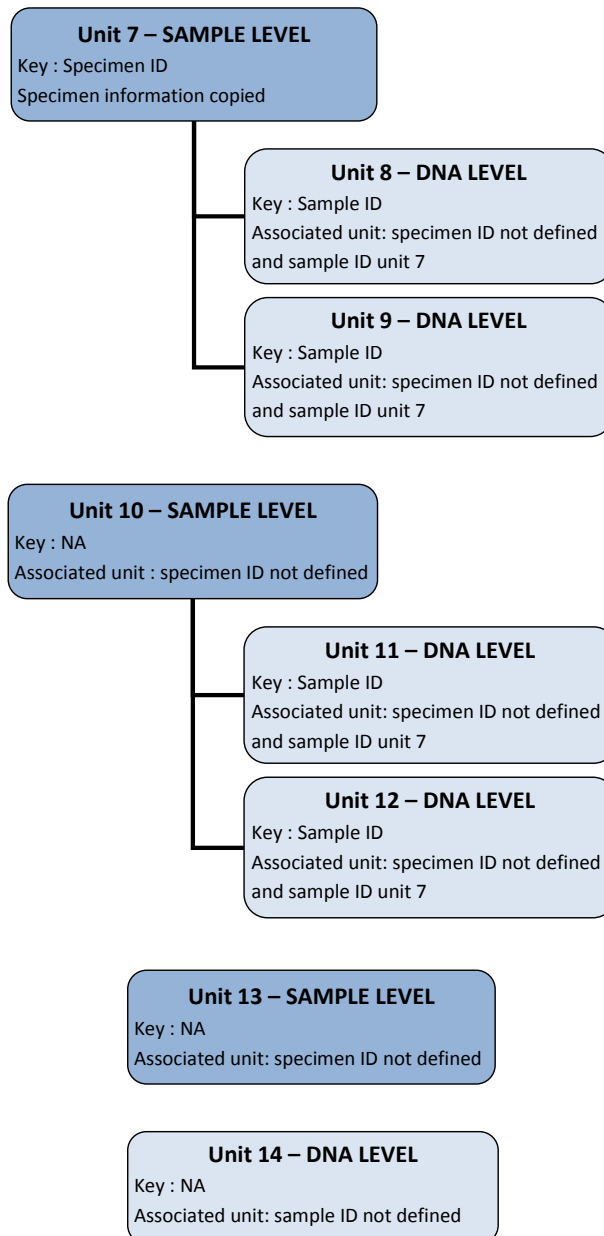


Figure 13 - Export XML file structure

## 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

In the SPECIMEN-sheet, supported fields are divided into 2 groups. The first group highlighted in dark blue corresponds to the supported values if you don't own the specimen. If you own the specimen, you can use all these fields as well as the second group highlighted in light blue.

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.

SPECIMEN				SAMPLE
OWNER?				
Yes		No		
Identification	Collecting event (continued)	Identification (additional info)	Properties (additional info)	sampleID
specimenID	collectionEndTimeH	additionalID	kindOfUnit	associatedSpecimenID
	collectionEndTimeM	code	totalNumber	sampleAcquiredFrom
Collecting event	expedition_project	accessionNumber	maleCount	partOfOrganism
samplingCode	samplingMethod	datasetName	femaleCount	sampleTissueType
ocean	fixation	isPhysical	sexUnknownCount	samplePreparationType
continent	ecology		socialStatus	samplePreservation
sea	siteProperty_1 to 10	Acquisition	specimenProperty_1 to 20	sample2Dbarcode
country	sitePropertyValue_1 to 10	acquisitionType	specimenPropertyValue_1 to 20	sampleInstitutionStorage
state_territory	localityNotes	acquiredFrom	hostClassis	sampleBuildingStorage
province		acquisitionDay	hostOrdo	sampleFloorStorage
region	Taxonomy	acquisitionMonth	hostFamilia	sampleRoomStorage
archipelago	classification	acquisitionYear	hostGenus	sampleColumnStorage
district	phylum		hostSpecies	sampleBoxStorage
county	classis	Taxonomy (additional info)	hostAuthor_year	sampleTubeStorage
department	ordo	identificationMethod	hostRemark	sampleNotes
island	superfamilia	identificationHistory		
city	familia	referenceString	Storage	
municipality	subfamilia	identificationNotes	institutionStorage	
naturalSite	genus	isPublished	buildingStorage	
populatedPlace	subgenus	publicationString	floorStorage	
exactSite	species	externalLink	roomStorage	
elevationInMeters	subspecies	urlPicture	laneStorage	
depthInMeters	author_year		columnStorage	
latitude	variety_form	Association	shelfStorage	
longitude	fullTaxonName	associatedUnitInstitution	container	
collectedBy	identifiedBy	associatedUnitCollection	containerType	
collectionStartDay	identificationDay	associatedUnitID	containerStorage	
collectionStartMonth	identificationMonth	associationType	subcontainer	
collectionStartYear	identificationYear		subcontainerType	
collectionStartTimeH			subcontainerStorage	
collectionStartTimeM	Properties		barcode	
collectionEndDay	statusType		conservation	
collectionEndMonth	sex			
collectionEndYear	lifeStage		Comments	
			notes	
				DNA
				dnaID
				dnaAdditionalID
				associatedSampleID
				dnaConcentration
				dnaAbsorbance260280
				dnaSize
				extractionTissue
				extractionMethod
				digestionTime
				digestionVolume
				elutionBuffer
				elutionVolume
				extractedBy
				extractionDay
				extractionMonth
				extractionYear
				dnaStorageMedium
				dna2Dbarcode
				dnaInstitutionStorage
				dnaBuildingStorage
				dnaFloorStorage
				dnaRoomStorage
				dnaFridgeOrDrawerStorage
				dnaBoxStorage
				dnaPositionStorage
				dnaNotes

## 2. List of fields by worksheet, with expected format, description and example(s)

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.

### a. SPECIMEN-sheet

Field	Format	Description	Example(s)
owner	X	Check this column if you own the specimen and wish to create it in DaRWIN	X or leave cell empty
specimenID		Unique identifier of specimen if exists. 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.	2013_Katanga_Termitidae_0001
additionalID		Additional identifier, for internal purpose only (classification of types, etc.)	Eutermes0001
code		A code associated to the specimen	RBINS/07-6-28
accessionNumber		IG Number for RBINS: institutional number given to each new group of items acquired by the institution and recorded in the collection registers	32327
datasetName		Name or code for the project, expedition, etc. for complementary information for the collection name. chosen in DaRWIN at the moment of the import	CongoRiverExp2013
isPhysical	Yes or no	Physical or observation	Yes
acquisitionType		Donation, purchase, etc.	Purchase
acquiredFrom		Former ownership (may be a person or an institution)	Example 1: RMCA; Example 2: Wallace Alfred (M.)
acquisitionDay	Numeric, 2 digits	Day of the acquisition date	8
acquisitionMonth	Numeric, 2 digits	Month of the acquisition date	10
acquisitionYear	Numeric, 4 digits	Year of the acquisition date	2007
samplingCode		A code that identifies all material or records resulting from a collecting or observation event	CRE201304
ocean		Ocean (administrative name)	Pacific Ocean
continent		Continent (administrative name)	Africa
sea		Sea (administrative name)	Red Sea
country		Country (administrative name)	Democratic Republic of the Congo
state_territory		State or territory, as a subdivision of a country (administrative name)	Katanga
province		Province (administrative name)	Bali
region		Region (administrative name)	Example 1: Upper Katanga; Example 2: Flemish Region
archipelago		Archipelago (administrative name)	Azores
district		District (administrative name)	Example 1: Zululand District; Example 2: North Somerset
county		County (administrative name)	Coconino County
department		Department (administrative name)	Loire-et-Cher
island		Island	Laing Island
city		Town, city, capital (administrative name)	Example 1: Cairns; Example 2: Kinshasa
municipality		Locality (administrative name) or urban administrative division	Likasi
populatedPlace		Populated place, village	Example 1: Tayabas; Example 2: Fortaleza
naturalSite		Natural site	Example 1: Parc National de la Salonga; Example 2: Great Coral Reef
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.	Example 1: Shituru Copper Mine; Example 2: 15km E of Tervuren along road N3
elevationInMeters	Numeric (unit: meters)	Altitude in meters	1020
depthInMeters	Numeric (unit: meters)	Depth in meters	20
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 or 10°58.88'S or -10.981333
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 or 26°44.20'E or 22.736666
collectedBy		Collector name (title)	Darwin Charles (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
collectionStartTimeH	Numeric, 2 digits, 24 hours format	Hours part of the time of day in hours from midnight for the starting of the collecting event - if not a time in hours, use the localityNotes field	10
collectionStartTimeM	Numeric, 2 digits, 24 hours format	Minutes part of the time of day in hours from midnight for the starting of the collecting event - if not a time in hours, use the localityNotes field	15

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
collectionEndTimeH	Numeric, 2 digits, 24 hours format	Hours part of the time of day in hours from midnight for the end of the collecting event - if not a time in hours, use the localityNotes field	17
collectionEndTimeM	Numeric, 2 digits, 24 hours format	Minutes part of the time of day in hours from midnight for the end of the collecting event - if not a time in hours, use the localityNotes field	5
expedition_project		Expedition or project name linked to the collecting event.	Recolte Institut/Fonds Leopold III 2013
samplingMethod		Technical means used for the sampling event	Example 1: Sticky trap and methyl eugenol; Example 2: net M5-F4
fixation		Fixative chemicals of the tissue or anesthetics used prior to tissue preservation	Anaesthesia
ecology		Biotope or any ecological information related to the collecting site	Example 1: Termite mount; Example 2: high tide; Example 3: surrounding vegetation
siteProperty_1		Observation or property of the collecting site (width, abundance, circumference, temperature, color, etc.)	Temperature AM
sitePropertyValue_1		Observation or property value of the collecting site (unit if applicable)	18°C
siteProperty_2		Observation or property of the collecting site (width, abundance, circumference, temperature, color, etc.)	Temperature PM
sitePropertyValue_2		Observation or property value of the collecting site (unit if applicable)	34°C
siteProperty_3		Observation or property of the collecting site (width, abundance, circumference, temperature, color, etc.)	Mount diameter
sitePropertyValue_3		Observation or property value of the collecting site (unit if applicable)	80 cm
siteProperty_4		Observation or property of the collecting site (width, abundance, circumference, temperature, color, etc.)	Mount height
sitePropertyValue_4		Observation or property value of the collecting site (unit if applicable)	124 cm
siteProperty_5		Observation or property of the collecting site (width, abundance, circumference, temperature, color, etc.)	Weather
sitePropertyValue_5		Observation or property value of the collecting site (unit if applicable)	Cloudy
siteProperty_6 to 10		Observation or property of the collecting site (width, abundance, circumference, temperature, color, etc.)	
sitePropertyValue_6 to 10		Observation or property value of the collecting site (unit if applicable)	
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
classification	Zoological/Botanical	Is it zoological or botanical?	Zoological
phylum		Phylum	Chordata
classis		Class	Actinopterygii
ordo		Order	Carnivora
superfamilia		Super family	Byrrhoidea
familia		Family	Termitidae
subfamilia		Subfamily	Macroroidinae
genus		Genus	Panthera
subgenus		Sub genus	Bryocamptus (Bryocamptus)
species		Species	tigris
author_year		Author, year or (Author, year) for the species	(Linnaeus, 1758) or Westwood, 1848
subspecies		Sub species	Example (Boa constrictor): imperator
variety_form		Variety or form	Example (Victorina steneles) : lavinia
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 : Panthera tigris (Mazak, 1968); Example 2 : Panthera tigris altaica Temminck, 1844; Example 3 : Papilio panthous var. numa Boisduval, 1836; Example 4 : Acmaeodera (Ptychomus) iridea Kerremans, 1902
identifiedBy		Identifier name (title)	Flemming Alexander (Dr.)
identificationDay	Numeric, 2 digits	Day of the identification date	2010
identificationMonth	Numeric, 2 digits	Month of the identification date	10
identificationYear	Numeric, 4 digits	Year of the identification date	12
identificationMethod		Material or method used for identification	Example 1: Dissecting microscope observation; Example 2: Molecular procedure
identificationHistory		Old genus or correction, comment on identification revision	Cucumaria (Semperia) bouvetensis Ludwig & Heding, 1935 (for Trachythone bouvetensis (Ludwig & Heding, 1935) which is the accepted name)
referenceString		References that were used by the identifier to provide the identification. Note that it is stored as a comment and NOT as a structured and searchable reference in DaRWIN.	World Register of Marine Species / Isopodes et Amphipodes de l'Expédition antarctique belge (s.y. Belgica). II. Bull. Mus. Hist. Nat., Paris 31: 296-299
publicationString		Published reference citing the unit. Note that it is stored as a comment and NOT as a structured and searchable reference in DaRWIN.	Rowson, Ben & Peter Tattersfield, 2013. Revision of Dadagulella gen. nov., the "Gulella radius group" (Gastropoda: Streptaxidae) of the eastern Afrotropics, including six new species and three new subspecies. European Journal of Taxonomy(37): 1-46.
identificationNotes		Additional information/remark about the taxonomy	Ask AR Wallace, for checking
urlPicture	url to a common location on the server	For one file, can be any type of file (image, pdf, txt, etc). files separated by a semicolon (;) - ex RBINS: smb://datastore/darwintmp/YOURFOLDER/yourimage.jpg; ask the IT team for a shared folder in your	smb://datastore/darwintmp/congriver2013/termitidae/00001.jpg
externalLink	url to an external website	External link where more information about the specimen can be found, links separated by a semicolon (;)	http://www.freshwaterbiodiversity.eu/
kindOfUnit		Part(s) of organism or class of materials represented by this unit (animal, mounted specimen, microscopic preparation, part of organism, etc.)	mounted, microscopic preparation, etc.

statusType		Type (paratype, holotype, etc.)	Paratype, holotype, lectotype, etc.
sex	Only: M, F, U, N or X	M=Male, F=Female, U=Unknown, N=Not applicable (for a fungus in a herbarium), X=Mixed (mixed specimens, males and females).	M
lifeStage		Phase or life stage	caterpillar, adult, larva, polype, etc.
socialStatus		Social status	workers, soliders, queen, etc.
totalNumber	Numeric	Total number of pieces that contains the record	12
maleCount	Numeric	Number of males in the record	3
femaleCount	Numeric	Number of females in the record	3
sexUnknownCount	Numeric	Number with unknown gender in the record	3
specimenProperty_1		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	widespan
specimenPropertyValue_1		Observation or property value (unit if applicable)	23 cm
specimenProperty_2		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	protonymphs count
specimenPropertyValue_2		Observation or property value (unit if applicable)	12
specimenProperty_3		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	original weight
specimenPropertyValue_3		Observation or property value (unit if applicable)	22 kg
specimenProperty_4		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	weight after dessication
specimenPropertyValue_4		Observation or property value (unit if applicable)	16 kg
specimenProperty_5		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	color
specimenPropertyValue_5		Observation or property value (unit if applicable)	yellow
specimenProperty_6		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	width
specimenPropertyValue_6		Observation or property value (unit if applicable)	23 cm
specimenProperty_7		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	antenna length
specimenPropertyValue_7		Observation or property value (unit if applicable)	17 mm
specimenProperty_8		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	Last erupted tooth
specimenPropertyValue_8		Observation or property value (unit if applicable)	M3
specimenProperty_9		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	larva count
specimenPropertyValue_9		Observation or property value (unit if applicable)	12
specimenProperty_10		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	circumference
specimenPropertyValue_10		Observation or property value (unit if applicable)	12 cm
specimenProperty_11 to 20		Observation or property of the item (weight, body measurements, winspan, last tooth, number of larva, pupa, juveniles, immatures, etc.)	
specimenPropertyValue_11 to 20		Observation or property value (unit if applicable)	
hostClassis		Host class	Mammalia
hostOrdo		Host order	Carnivora
hostFamilia		Host family	Canidae
hostGenus		Host genus	Canis
hostSpecies		Host species	lupus
hostAuthor_year		Author, year or (Author, year) for the host taxon name	Linnaeus, 1758
hostRemark		Additional information/remark about the host	Not collected
associatedUnitInstitution		Institution ownership of the associated unit	RMCA
associatedUnitCollection		Associated unit collection name	Subsampling_Eutermes
associatedUnitID		Associated unit ID	Macrotermes_0001
associationType		Association type (DNA extract, slide, host, etc.)	Example 1: blood; Example 2: fin clip; Example 3: feather
barcode		2DBarcode	AB42715507
conservation		Fixative used for long term conservation	Example 1: Ethanol; Example 2: silica gel beds
institutionStorage		Institution storage (acronym or full name)	RBINS
buildingStorage		Building storage	De Vestel
floorStorage		Floor storage	6
roomStorage		Room storage	28
laneStorage		Lane storage in the room	2
columnStorage		Column or cupboard storage in the lane	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	A1
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



## b. SAMPLE-sheet

Field	Format	Description	Example(s)
sampleID		Unique identifier of sample if exists. Example of format: [YYYY]_[CollectionOrDatasetCode]_[SubGroup]_[Iterative_nb]. The collection code may include an acronym representing the expedition and/or the institutional registration number. The subgroup may be the order/family concerned. It can be the ID of the specimen with an iterative number as suffix.	2013_Katanga_Termitidae_0001-sample01
associatedSpecimenID		Specimen ID associated to this sample	2013_Katanga_Termitidae_0001
sampleDonorInstitution		Institution or person that donates this sample	Example 1: Carson Rachel (Mrs.) ; Example 2: Australian National Wildlife Collection
partOfOrganism		Part of organism that furnished this sample	Example 1: fin clip; Example 2: liver; Example 3: heart
sampleTissueType		Tissue type used for extraction	Example 1: muscle; Example 2: blood
samplePreparationType		Preparation of this sample	Field or lab sampling, sterilisation method
samplePreservation		Information about the preservation of this sample	ethanol -20°C
sample2Dbarcode		2D barcode	AB42715495
sampleInstitutionStorage		Institution storage (acronym or full name)	RMCA
sampleBuildingStorage		Building storage	CAPA
sampleFloorStorage		Floor storage	-1
sampleRoomStorage		Room storage	Conservatory A
sampleColumnStorage		Column or cupboard storage	3
sampleBoxStorage		Box storage	12
sampleTubeStorage		Tube storage	A1
sampleNotes		Additional information/remark about the sample/tissue that doesn't fit elsewhere	

## c. DNA-sheet

Field	Format	Description	Example(s)
dnalD		Unique identifier of DNA-extract if exists. Example of format: [YYYY]_[CollectionOrDatasetCode]_[SubGroup]_[Iterative_nb]. The collection code may include an acronym representing the expedition and/or the institutional registration number. The subgroup may be the order/family concerned. It can be the ID of the sample with "DNA" and an iterative number as suffix.	2013_Katanga_Termitidae_0001-sample01-DNA01
dnaAdditionalID		Additional identifier, for internal purpose only (accession number in BOLD systems, etc.)	AFNF001-12
associatedSampleID		Sample ID associated to this DNA-extract	2013_Katanga_Termitidae_0001-sample01
dnaConcentration	Numeric	DNA concentration (ng/μl)	29.47
dnaAbsorbance260280	Numeric	Ratio of absorbance at 260 and 280 nm	2.47
dnaSize	Integer	Length of the DNA-extract (bp)	300
extractionTissue		Type of tissue used for extraction	Example 1: liver; Example 2: soft tissue
extractionMethod		DNA Extraction protocol	Example 1: NS-tissue; Example 2: Qiagen DNeasy Blood & Tissue Kit
digestionTime		Digestion time	Example 1: 2h; Example 2: overnight
digestionVolume	Numeric	Digestion volume (μl)	100
elutionBuffer		Elution buffer	Example 1: Tris; Example 2: AE
elutionVolume		Elution volume (μl)	Example 1: 200; Example 2: 1 x 100
extractedBy		Extraction operator	Watson James
extractionDay	Numeric, 2 digits - full date required	Day of extraction	12
extractionMonth	Numeric, 2 digits - full date required	Month of extraction	1
extractionYear	Numeric, 4 digits - full date required	Year of extraction	2013
genBank		GenBank number	Example Protein: AGB95530; Example nucleotide: A23456
dnaStorageMedium	dry or frozen	Storage medium for the DNA-extract: dry or frozen	dry
dna2Dbarcode		Rack storage	AB43158587
dnaInstitutionStorage		Position	RBINS
dnaBuildingStorage		2D barcode	De Vestel
dnaFloorStorage		Floor storage	4
dnaRoomStorage		Room storage	JEMU office
dnaFridgeOrDrawerStorage		Fridge or drawer storage	Genvault1
dnaBoxStorage		Box storage	2
dnaPositionStorage		Position storage	A1
dnaNotes		General comment about the DNA-extract	Example 1: Working Solution - in solution 33.4μl; Example 2: problem during extraction

### 3. Link between the fieldsheet and the template

Through the development of this template, we tried to develop standardized data format fitting to our collection management system. In parallel, we prepared a standard fieldsheet in which every field has a corresponding one in the template.

Template field	Fieldsheet (page 1)
recordID	Unique ID
expedition_project	Project or expedition name
collectedBy	Collector
conservation	Preservation
barcode	Barcode
Collecting event	
samplingCode	Collecting event code
Collection date	
collectionStartDay	Start date (YYYY-MM-DD)
collectionStartMonth	
collectionStartYear	
collectionStartTimeH	Start time (HH:MM)
collectionStartTimeM	
collectionEndDay	End date (YYYY-MM-DD)
collectionEndMonth	
collectionEndYear	
collectionEndTimeH	End time (HH:MM)
collectionEndTimeM	
Sampling location	
ocean	Ocean
continent	Sea
sea	Continent
country	Country
state_territory	State or territory
province	Province
region	Region
archipelago	Archipelago
district	District
county	County
department	Department
island	Island
city	City
municipality	Municipality
naturalSite	Natural site
populatedPlace	Populated place
exactSite	Exact site
elevationInMeters	Altitude (m)
depthInMeters	Depth (m)
latitude	Latitude (preferred format: decimal degrees)
longitude	Longitude (preferred format: decimal degrees)
ecology	Ecology
Collection strategy	
samplingMethod	Sampling method
fixation	Fixatives or anaesthetics
Site properties or observations	
siteProperty_1 to 10	Property or observation
sitePropertyValue_1 to 10	Value
Notes	
localityNotes	Notes
General Notes	
notes	General Notes

Template field	Fieldsheet (page 2)
<b>Classification</b>	
<b>Taxon name</b>	
species	Species
subspecies	Subspecies
author_year	Author, year
<b>Complete taxonomy</b>	
phylum	Phylum
classis	Class
ordo	Order
superfamilia	Superfamily
familia	Family
subfamilia	Subfamily
genus	Genus
subgenus	Subgenus
<b>Additional information</b>	
variety_form	Variety or form
commonName	Common name
identificationMethod	Identification method
identifiedBy	Identifier
identificationDay	Identification date
identificationMonth	
identificationYear	
identificationNotes	Notes
<b>Specimen properties or observations</b>	
<b>General information</b>	
kindOfUnit	Tissue/part of organisme
sex	Sex
lifeStage	Life stage
totalNumber	Total number
maleCount	Number of males
femaleCount	Number of females
sexUnknownCount	Number of unknown sex
<b>Additional information</b>	
specimenProperty_1 to 20	Property or observation
specimenPropertyValue_1 to 20	Value
<b>Host</b>	
hostClassis	Class
hostOrdo	Order
hostFamilia	Family
hostGenus	Genus
hostSpecies	Species
hostAuthor_year	Author, year
hostRemark	Notes

# FIELD SHEET

Unique ID	<input type="text"/>	Project or expedition name	<input type="text"/>
Collector	<input type="text"/>		

Barcode	<input type="text"/>	Preservation	<input type="text"/>
---------	----------------------	--------------	----------------------

<b>COLLECTING EVENT</b>		Collecting event code <input type="text"/>	
Collection date			
Start date (YYYY-MM-DD)	Start time (HH:MM)	End date (YYYY-MM-DD)	End time (HH:MM)
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Sampling location			
Ocean	Continent	Sea	
<input type="text"/>	<input type="text"/>	<input type="text"/>	
Country	State or territory	Province	
<input type="text"/>	<input type="text"/>	<input type="text"/>	
Region	Archipelago	District	
<input type="text"/>	<input type="text"/>	<input type="text"/>	
County	Department	Island	
<input type="text"/>	<input type="text"/>	<input type="text"/>	
City	Municipality		
<input type="text"/>	<input type="text"/>		
Natural site	Populated place	Exact site	
<input type="text"/>	<input type="text"/>	<input type="text"/>	
Altitude (m)	Depth (m)	Ecology	
<input type="text"/>	<input type="text"/>	<input type="text"/>	
Latitude (preferred format: decimal degrees)	Longitude (preferred format: decimal degrees)		
<input type="text"/>	<input type="text"/>		
Collecting strategy			
Sampling method			
<input type="text"/>			
Fixatives or anaesthetics <input type="text"/>			
Site properties or observations			
Property or observation	Value	Property or observation	Value
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Notes <input type="text"/>			

<b>GENERAL NOTES</b>
<input type="text"/>

**CLASSIFICATION****Taxon name**

Species

Subspecies

Author, year

**Complete taxonomy**

Phylum

Class

Order

Superfamily

Family

Subfamily

Genus

Subgenus

**Additional information**

Variety or form

Common name

Identification method

Identifier

Identification date

Notes

**SPECIMEN PROPERTIES OR OBSERVATIONS****General information**

Tissue/part of organism

Sex

Life stage

Total number

Number of males

Number of females

Number of unknown sex

**Additional information**

Property or observation

Value

Property or observation

Value

**HOST**

Class

Order

Family

Genus

Species

Author, year

Notes

#### 4. Link between the template and the BOLD Systems Submission format

New specimen submission to BOLD can be made with the Data Submission Template spreadsheet, available at [http://www.boldsystems.org/index.php/resources/handbook?chapter=3\\_submissions.html](http://www.boldsystems.org/index.php/resources/handbook?chapter=3_submissions.html). This data spreadsheet consists of 4 worksheets: a main specimen identifier worksheet (voucher info) that is linked to three other worksheets (taxonomy, specimen details and collection data).

The minimal requirements for a new specimen record on BOLD are: Sample ID, Field ID and/or Museum ID, Institution Storing, Phylum and Country

The information that you store in the template can be used in the BOLD Data Submission spreadsheet. The following table gives you the correspondence between the fields present in the template and the ones you can find in the BOLD Data Submission spreadsheet. The colour in the columns for BOLD fields is the same as in the BOLD Data Submission Template.

BOLD systems		ABCD Template	
BOLD field	BOLD sheet	ABCD Template field	ABCD Template sheet
Sample ID	Voucher Info	sampleID	SAMPLE
Museum ID	Voucher Info	specimenID	SPECIMEN
Collection Code	Voucher Info	datasetName	SPECIMEN
Institution Storing	Voucher Info	institutionStorage	SPECIMEN
Collectors	Collection Data	collectedBy	SPECIMEN
Collection Date	Collection Data	collectionStartDay	SPECIMEN
		collectionStartMonth	SPECIMEN
		collectionStartYear	SPECIMEN
Country/Ocean	Collection Data	ocean	SPECIMEN
Country/Ocean	Collection Data	country	SPECIMEN
State/Province	Collection Data	state_province	SPECIMEN
Region	Collection Data	region_district	SPECIMEN
Exact Site	Collection Data	exactSite	SPECIMEN
Latitude	Collection Data	latitude	SPECIMEN
Longitude	Collection Data	longitude	SPECIMEN
Elevation	Collection Data	elevationInMeters	SPECIMEN
Depth	Collection Data	depthInMeters	SPECIMEN
Event Time	Collection Data	collectionStartTimeH	SPECIMEN
		collectionStartTimeM	SPECIMEN
Habitat	Collection Data	ecology	SPECIMEN
Sampling Protocol	Collection Data	samplingMethod	SPECIMEN
Collection Notes	Collection Data	localityNotes	SPECIMEN
Collection Event ID	Collection Data	samplingCode	SPECIMEN
Phylum	Taxonomy	phylum	SPECIMEN
Class	Taxonomy	classis	SPECIMEN
Order	Taxonomy	ordo	SPECIMEN
Family	Taxonomy	familia	SPECIMEN
Subfamily	Taxonomy	subfamilia	SPECIMEN
Genus	Taxonomy	genus	SPECIMEN
Species	Taxonomy	species	SPECIMEN
Identifier	Taxonomy	identifiedBy	SPECIMEN
Identification Method	Taxonomy	identificationMethod	SPECIMEN
Taxonomy Notes	Taxonomy	identificationNotes	SPECIMEN
Sex*	Specimen Details	sex	SPECIMEN
Life Stage	Specimen Details	lifeStage	SPECIMEN
Notes	Specimen Details	notes	SPECIMEN
Tissue Descriptor	Specimen Details	kindOfUnit	SPECIMEN

\*In BOLD : Male/female/hermaphrodite only