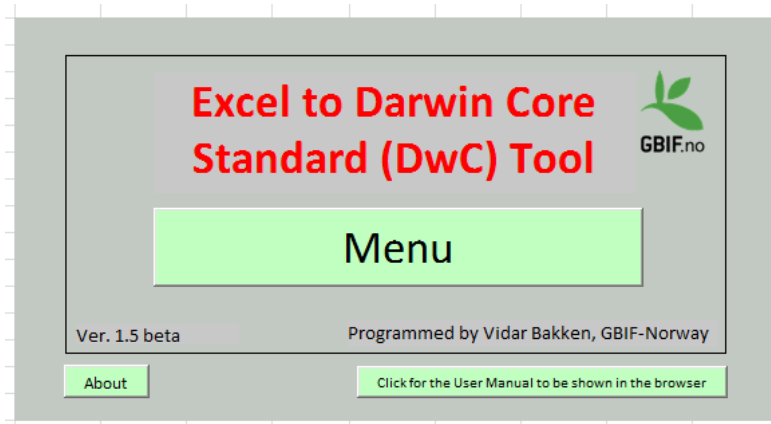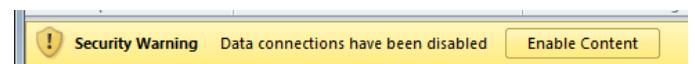# Excel to Darwin Core Standard (DwC) Tool

Vidar Bakken (GBIF-Norway)

This is a macro excel spreadsheet for setup, data entry, data validation and file export to DwC files formatted for the GBIF IPT, the Integrated Publishing Toolkit (https://www.gbif.org/ipt). With the spreadsheet you can register Sampling-event (https://www.gbif.org/sampling-event-data) and Occurrence datasets (https://www.gbif.org/data-quality-requirements-occurrences). It is also possible to attach MeasurementsOrFacts (MoF) (https://tools.gbif.org/dwca-validator/extension.do?id=dwc:MeasurementOrFact), Extended MeasurementsOrFacts (EMoF) and Simple Multimedia (https://tools.gbif.org/dwca-validator/extension.do?id=gbif:Multimedia).

As the spreadsheet contains macros programmed in Visual Basic for Applications (VBA) it may be a problem to send the file as an attachment to an email. The extension of excel macro files are *.xlsm or *.xlsb. Such files are considered as dangerous by anti-virus programs and are often stopped. A possibility is to zip the file before sending it as an attachment. A better way to distribute such files are by Dropbox, OneDrive or other file hosting services. VBA is built into most Microsoft Office applications, including Office for Mac OS X (except version 2008).
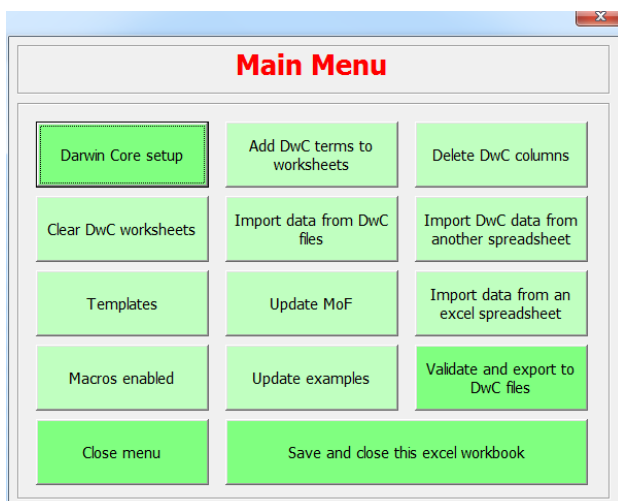
When you start the spreadsheet for the first time you must probably enable the use of the macros. Often you must click on a message box under the top menu in excel:

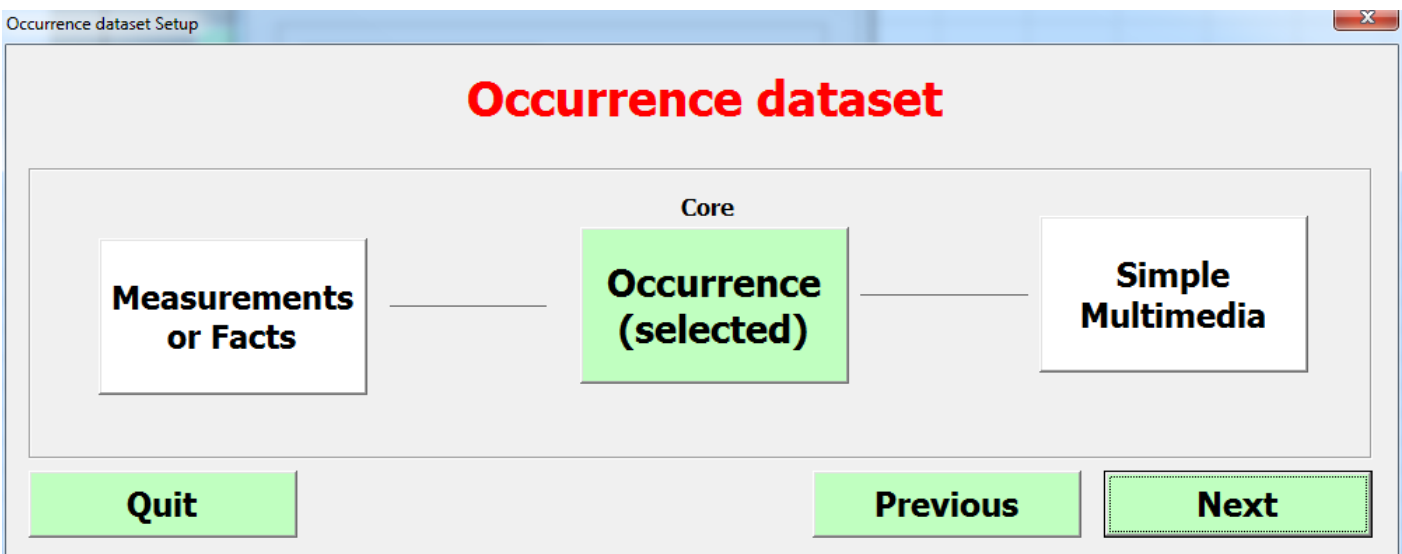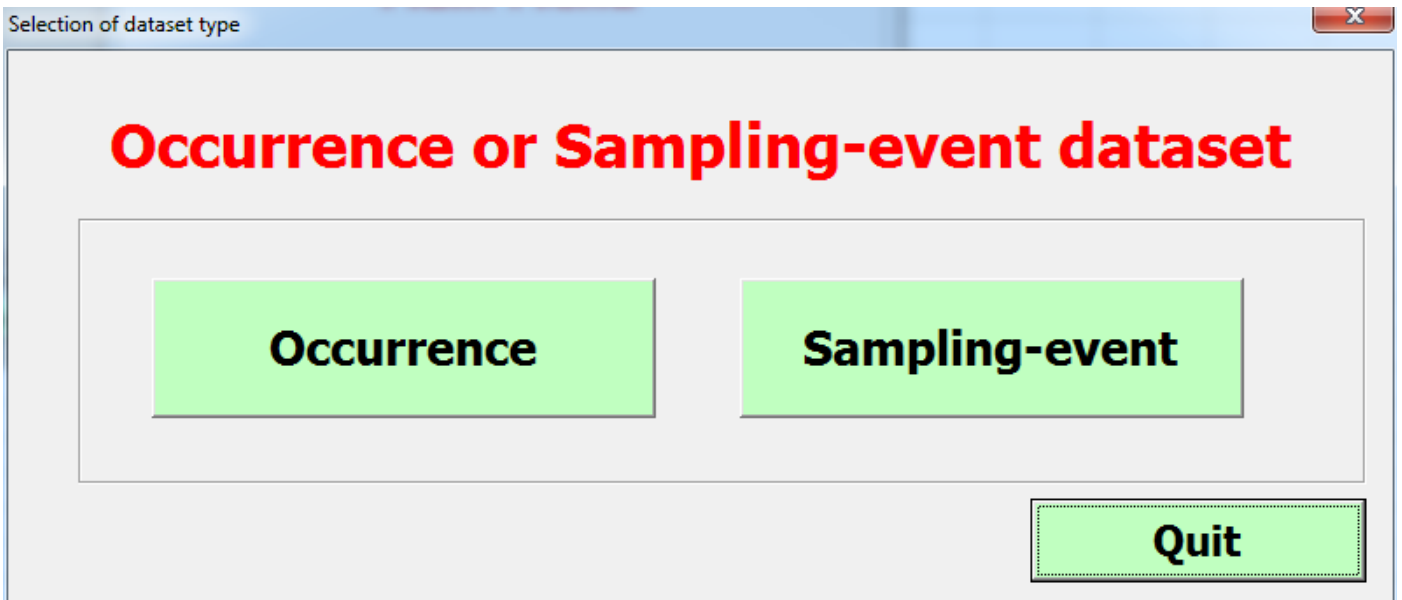*Main menu in the spreadsheet*

Get started

Click «Menu» on the Administration sheet and you will see a main menu (left).

The first task is to format the DwC files to be used for the data entry. Click «Darwin Core Setup». Then you select if you want to define an occurrence or a sampling-event dataset (see

Selection of occurrence or sampling-event dataset. If you select an occurrence dataset you can also attach the extensions Measurements or Facts and/or Simple Multimedia.

links above).

The simplest format is an occurrence dataset. Click «Occurrence» and you can define what extensions you want to connect to the occurrence table (core), i.e. Measurements or Facts and/or Simple Multimedia. Click the buttons to select.

Click «Next» to select the DwC terms to be used in the Occurrence sheet. A set of terms is already proposed. You can delete or add terms to this list. Also the order of terms can be altered by clicking the buttons marked «Up» and «Down». By selecting a term the description will be shown in the textbox to the left. You can aslo display the terms grouped in the combobox below the term names (Record-level, Occurrence, Event etc.).

Click next to proceed. The next step will depend on what other extensions which are selected in the setup. If you marked «Measurements or Facts» a new form will be visible.

Give the terms you want to include (measurementType and Unit). Click «Register» and the selection is shown in the listbox. You can add as many terms as you want. When completed, click «Next».

If you have selected Simple Multimedia a new form will be visible. This is identical to the form with Occurrence terms, but includes terms for media only. As before you can delete, add and alter the order of terms. By clicking «Next» you have ended the session for the setup
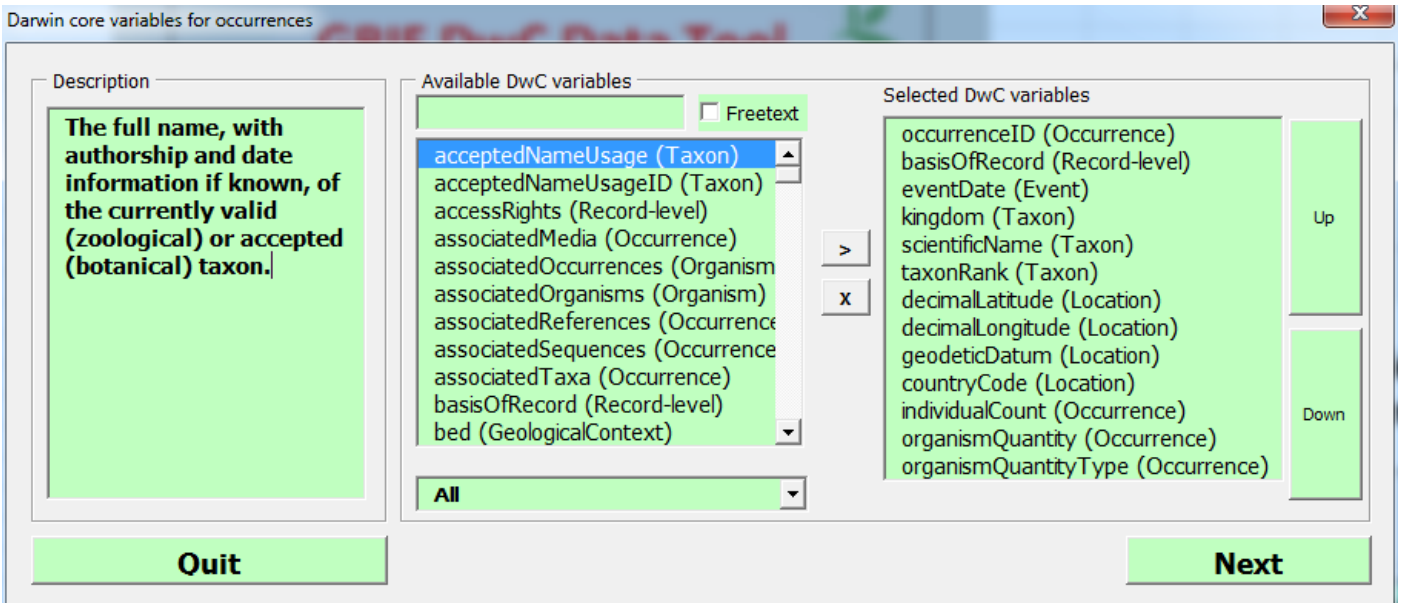
Selection of terms in an occurrence dataset.

You can define as many terms as you want in the extensions Measurements or Facts and Extended Measurements or Facts. Remember to click «Register» after each new term and check that they are visible in the list above. You can delete terms from the list by right-click on the desired term.
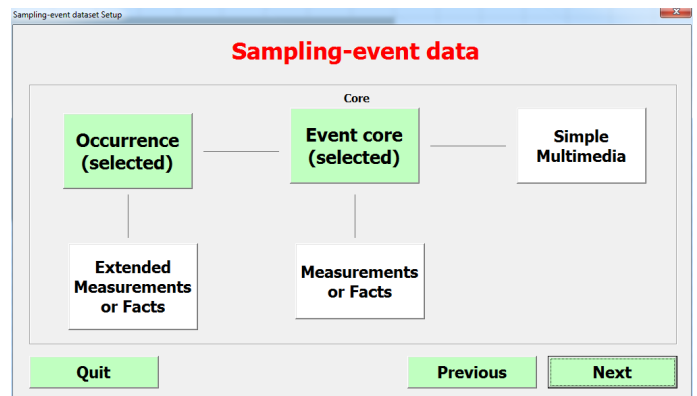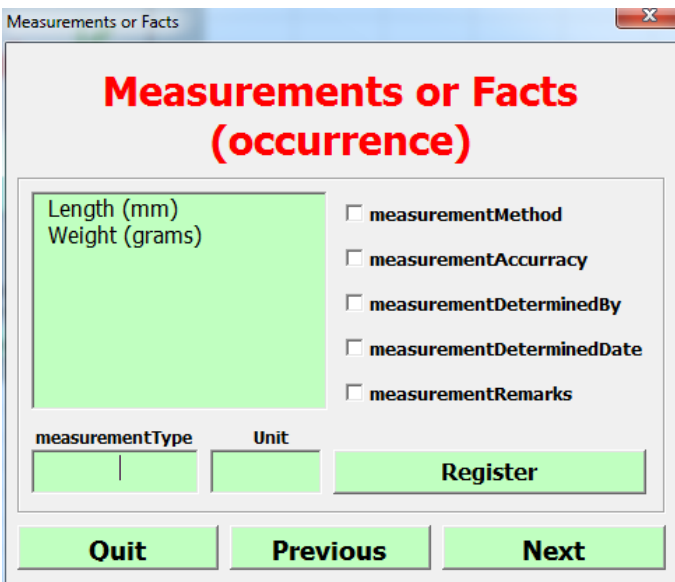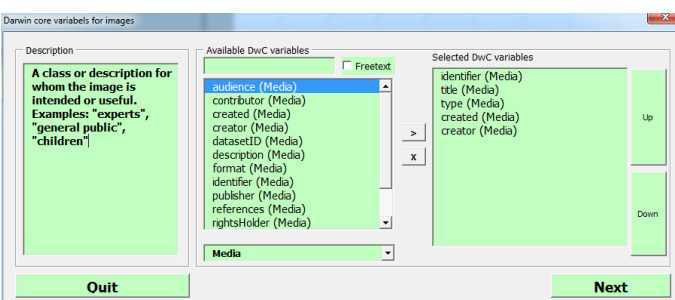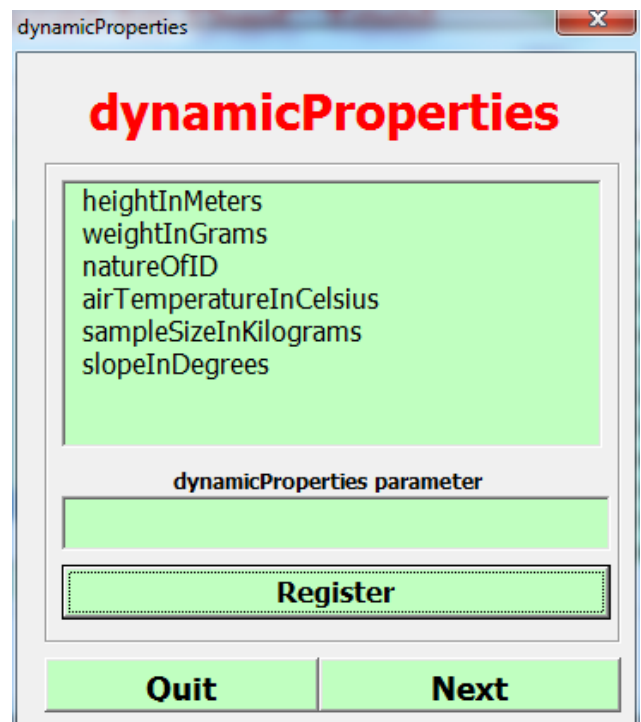

If you select a Sampling-event dataset you can also attach the extensions Measurements or Facts, Extended Measurements or Facts and/or Simple Multimedia.
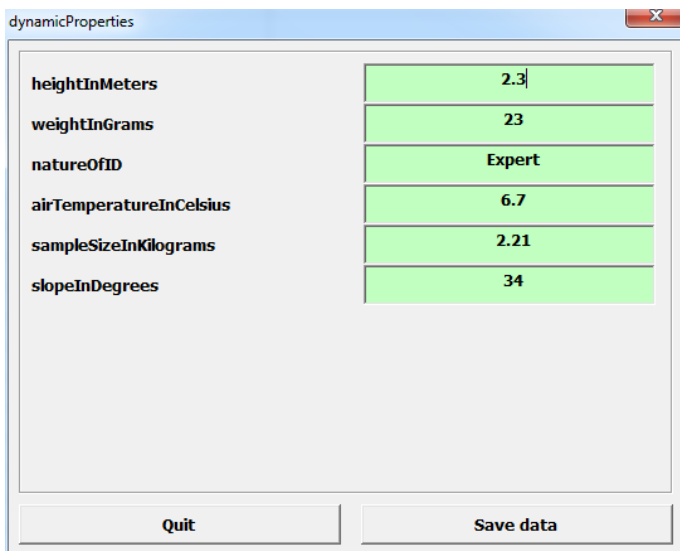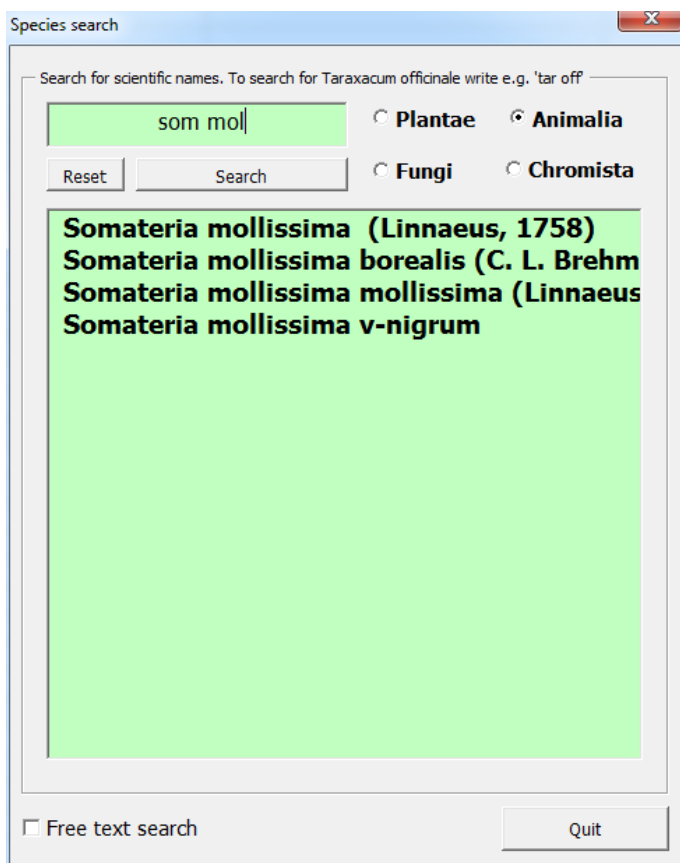

DwC terms setup for Simple Multimedia.

of an Occurrence dataset with the connected extensions.

If you use dynamicProperties you have the possibility to define up to 10 dynamicProperties terms.

By double-clicking in a dynamicProperty cell you can register the values in a form. The data is saved in json format.

The search for species names. The systematic lists are from the Norwegian Biodiversity Information Centre.

The setup for a Sampling-event dataset is quite similar to the Occurrence dataset, but here you must also define terms for the event table. In addition you can also have Extended Measurements of Facts connected to the Occurrence table. Otherwise the procedure is the same.
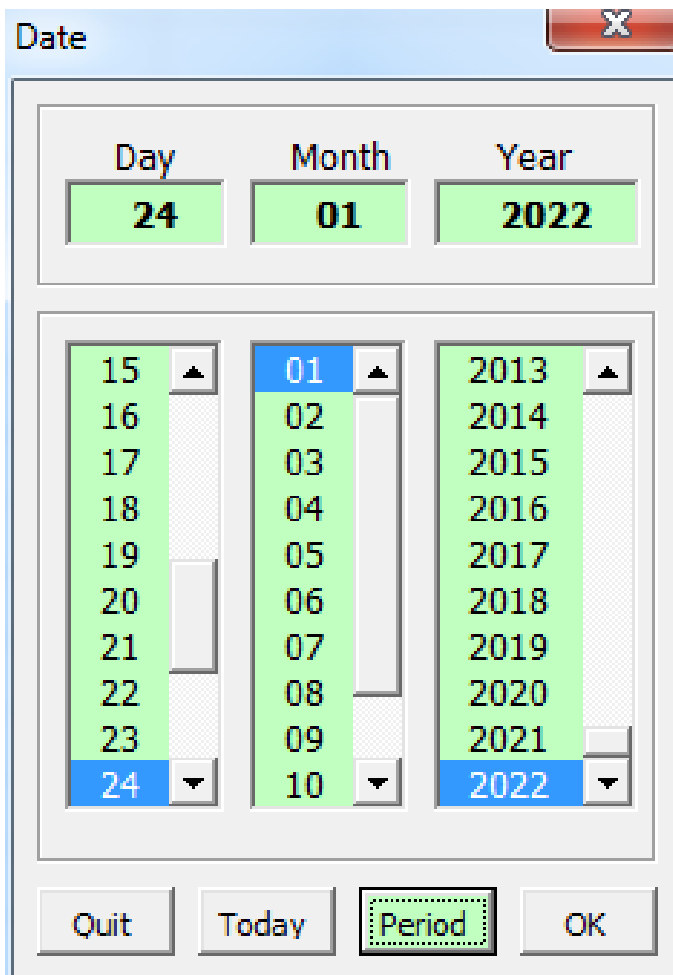
Date selection. By clicking «Period» you can register a date period (from - to).

Data entry

From the main menu click «Close menu» to return to the spreadsheet. Dependent on your setup different sheets will be visible («Events», «Occurrences», «MeasurementsorFacts», «ExtendedMeasurementsorFacts» and/or «SimpleMultimedia».

The terms in the top row in each sheet will be same as you selected and ordered during the setup.

If you selected an Occurrence dataset you must start in the Occurrence sheet. By clicking at the names in the top row a message box will appear with the description of the DwC term.

Start with «basisOfRecord». Double click in the first cell and select from the list. As you select an UUID will be placed in the first column (occurrenceID). In general you have not access to cells with UUIDs. The reason is that they should not be altered. Continue with the next column and double click in the cell.

(2) Click on the *place name* or *coordinates*. By clicking the *place name* you can also transfer this name if you have used the term 'locality' in DwC. Be aware that the waypoint may be changed if you select *place name*. Click *coordinates* if you want the most accurate coordinates and fill out the locality name yourself.

(1) Click at the point where you want to register the coordinates to make a waypoint.

(3) Finally you click this button to transfer the coordinates/place name to the sheet.

Våler i Østfold
1592
59.501843, 10.906098

Quit without position

Return to worksheet with marked position

Procedure to transfer coordinates and place names from Google Maps to the sheet by using the integrated form in the spreadsheet. You activate the form by double-clicking in a cell in the columns «decimalLatitude» or «decimalLongitude».

If there are available examples for input they will be selectable from a list. If no examples are available a keyboard will appear. Use this or write directly in the cell.

By double clicking in a cell under scientificName you will get a form where you can search for species. Select Plantae, Animalia. Fungi or Chromista. The most effective way to search for a species is to give the first letters in the genus and species name (scientific). E.g you want to search for Common Eider (*Somateria molissima*) you can e.g. enter «som mol» as in the example. It is also a possibility to do a free text search by clicking the check box in the bottom. If you write the scientific names yourself or paste in a list you can also validate the data in the list. Click the cell on the top "scientificName (Taxon)". After the display of the term description a form is displayed (below). Select the group(s) in the list and click "Yes". Names not validated are shown in red. If you will give the author you must put it in brackets "[]".
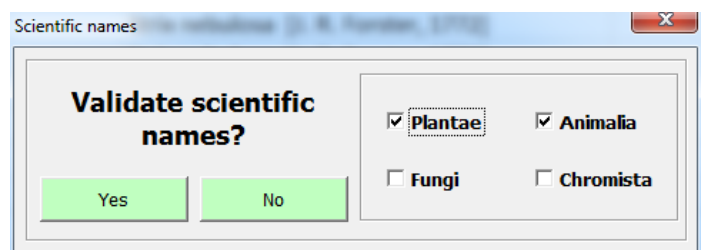

Validation of scientific names. Select the group(s) and click "Yes". Do not select other groups than those that are in the list. More groups will delay the validation.

The systematic register used is from the Norwegian Biodiversity Information Centre. It mostly covers species occurring in Norway/Svalbard/Jan Mayen, but includes also many species living elsewhere. If you don't find the species you are looking for you can write the full species name in the cell yourself.

By double clicking in the column with date you can select the date from a form. The coordinates can be fetched from Google Maps. By double clicking in the columns «decimalLatitude» or «decimalLongitude» you can display Google Maps in a form. Search your target area and mark a waypoint on the actual site. Remember to click the coordinates given in the bottom of the form (see figure). There is also an option for transfer of the locality name. Then click «Return to worksheet with marked position». The coordinates in decimal degrees are placed in the columns for coordinates. The locality name can be transfered if you have the term «locality» defined. See the figure for details.

You will find the Measurement or Facts terms after the occurrence terms in the top column in the sheet. An example could be «MoF (Weight grams)», i.e. the body mass in grams. Enter a value in a cell and the cell turns from red to green. If you have selected more terms than measurementType and Unit you will be directed to MeasurementotFacts sheet. Fill in the extra information and return to the Occurrences sheet by double click in any cell in the same row. Do not edit the UUID (occurrenceID) which is transferred from the Occurrence sheet to the MeasurementsorFacts sheet. Continue to register in the occurrence sheet. The next time you register the same MoF term you don't need to be transferred to the MeasurementorFacts sheet as the extra terms are copied from the previous registration of the same term. If needed you can double click in the cell to have the possibility to edit the data. It is also possible bo select the MeasurementorFacts sheet manually and do the edits.

If you have selected Simple Multimedia you will find a column named «Images». Double
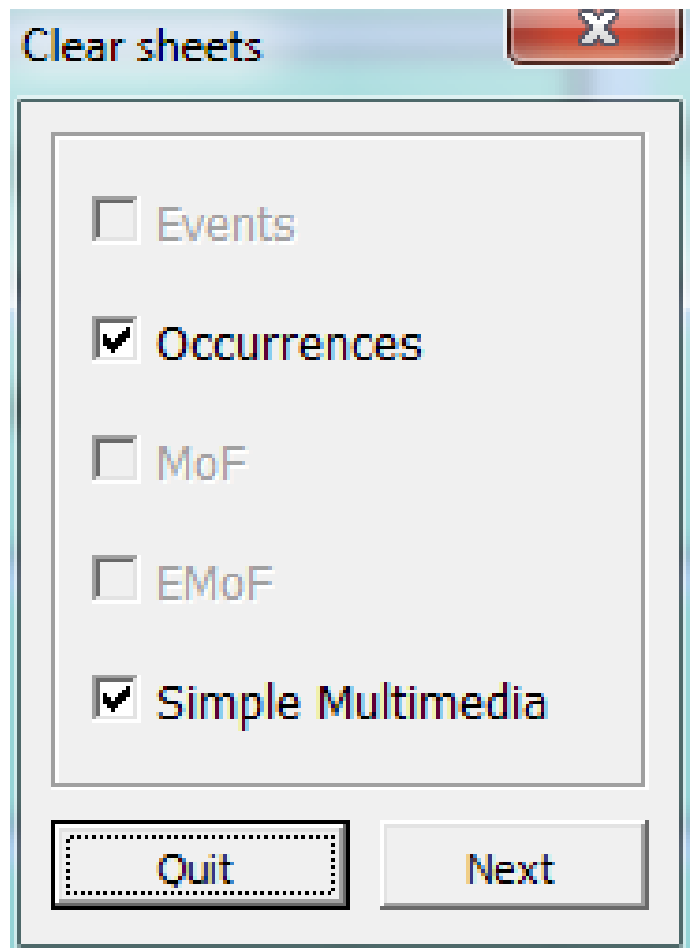
Select the sheets you want to clear. It is possible to only clear the data or also the DwC terms.

click in a cell in this column and you are directed to the SimpleMultimedia sheet. Fill in the values and return to the Occurrence sheet by double click in any cell in this row. Then «1 image» will be written in a green cell. You can connect as many images as you want to the same occurrence record.

If you select to generate a Sampling-event dataset you have the possibility to connect an Extended MeasurementorFact sheet to the Occurrence table. The ordinary MeasurementorFact and the SimpleMultimedia tables will be connected to the Event table as is the core in the dataset.

For a Sampling-event dataset you start to enter data in the Event table, then the Occurrence table. The term «eventID» is in both tables. To connect each row in the Occurrence table to the Events, click in a cell under «eventID» in the Occurrence sheet. Then you get a list of all registered events. Select one to connect the event and occurrence table. Alle rows in the

Occurrence sheet must be connected to a row in the Event table.

Validate and export to DwC

Start the main menu and select «Validate and export to DwC export files». Confirm and the files are written to the same folder where you have the excel spreadsheet.

The data will be validated and eventually problems will be written to a list which is displayed. Correct all mistakes before you export the data to the final text files. The validation is not so detailed as in the IPT. E.g. coordinates are not be validated in relation to the country, stateprovince or the county given.

The files are named according to the data type (Events, Occurrences, MoF, EMoF and Images). They also get a date stamp as part of the file name (e.g. Occurrences_20211215_121510. txt). The format is tab-text files.

Clear DwC worksheets

This function gives the possibility to clear the data in the sheets. Select the sheets and if you want to clear just the data or also the DwC
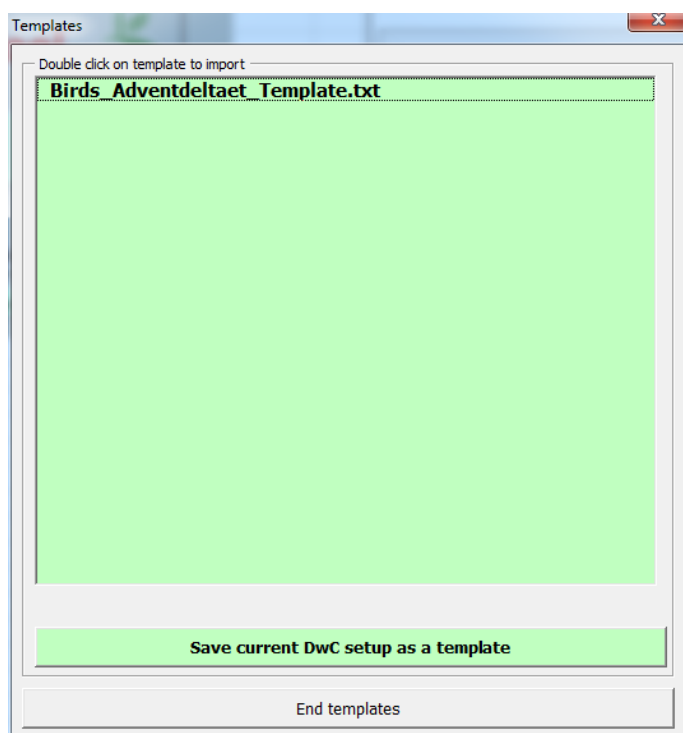
You can save the registered DwC term as templates. Later they are available for import and will be visible in the list. No data is saved in a template, terms only.

setup (top row).

Add DwC terms to the worksheets

If you need more DwC terms than you initially selected you can add others in this menu. It works similar to the first setup of terms.

Import data from DwC files

DwC files in the same format as you export can also be imported. Before the import the sheets are cleared. You can select which type of data you want to import. Follow the guidelines from the program.

Import DwC data from another spreadsheet

DwC data from another spreadsheet in the same format can be imported. This is an useful function when e.g. you need to transfer data to a newer version of the spreadsheet.

Import DwC data from an excel spreadsheet

DwC data from an excel spreadsheet can be imported from both sample-event and occurrence datasets. In a Sample-event dataset the first column must have the heading "eventID". In the occurrence sheet the first two columns must be "eventID" and "occurrenceID".

Select the excel file and there is a check of the terms. Terms not automatically mapped to a valid DcW term must be mapped or skipped manually. Such terms are listed in the upper list (right). You map the terms by selecting the correct term in the list to the left and click "Map". Then the term is written to list in the bottom. When you import a sample-event dataset you must do one mapping for each worksheet. Columns not suitable for the DwC format can be skipped. Select the term and click 'X'. The term is written in the bottom list marked as "Skipped".

If there are no UUIDs for the rows it is
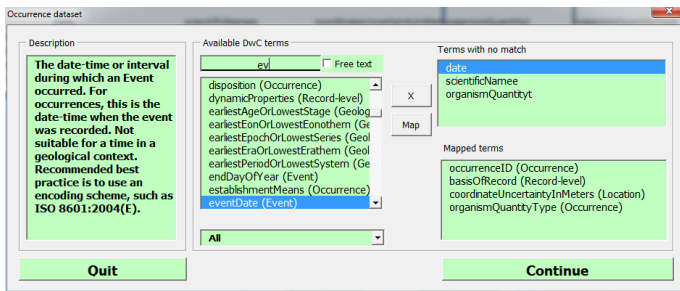
When migrating data from excel you can map terms not found in the DwC vocabulary (listed in the upper right list). Mark the unmapped term and the correct term in the DwC list. Click "Map" and the term is written to the list in the bottom. It is also possible to skip columns not suitable for the DwC format. Select the term and click 'X'. The term is written to the list in the bottom as "Skipped".

possible that the program automatically gives an UUID to each row (eventID in a sample-event dataset and occurrenceID in an occurrence dataset.

Update MoF

This function updates Measurements or Facts (incl. Extended Measurements or Facts) from the worksheets Events and Occurrences. If you enter data directly in the cells in the Occurrences and Events sheets the data is automatically updated in the Measurements or Facts worksheets.

This function is useful if you e.g. paste in data from another data source. Then the update is not automatic, and you can do the update by using this function. The function is also automatically activated during the data validation process in order to make sure that all links between Occurrences/Events and Measurements or Facts are updated.

Templates

It is possible to save the DwC setup as templates. Click «Templates» in the main menu. Click «Save current DwC setup as template». You can give it a name before it is saved as a text file in the program folder. The extension «_Template. txt» will be added to the name. The template will be visible in the list above.

Such templates can be reused by all users of the spreadsheet. Just copy the file to the program folder and the template will be available.

Delete DwC columns

You can delete columns in all DwC sheets by clicking «Delete columns» in the main menu. Follow the instructions given. If you have entered data in these columns all these data will be deleted.

Macros

By default the macros are enabled. By clicking «Macros enabled» you can disable the macros in the spreadsheet. This is practical if you want to edit cells that otherwise are closed. Remember to click the button once more to enable the macros when completed.

Examples

By double-clicking in a cell in the different worksheets you can get examples of input values for terms. It is possible to add more examples to these lists. In the worksheet «Examples» terms are listed with the associated examples. After you have entered the data you can update the example list with the data in the tables. Click «Update examples» in the main menu. The program checks all data that are entered and you get the possibility to update the example lists with new registrations. You can also add them manually if you want them to be visible on the lists from the start.

It is also possible to add new terms with examples. Make sure that you write the term name in accordance with the GBIF standard. Register the new terms before you start the update, or register the examples yourself.

Summary

This is a beta version of the program (ver. 1.5 beta). It is quite possible that errors may occur. Please send me a message if you experience any problems or if you have proposals for improvements.

Good luck with your GBIF and DwC work!

Questions and comments can de directed to:
Vidar Bakken
GBIF-Norway
E-mail: vidar.bakken@usit.uio.no