VertNet's Migrator Toolkit

Instructions for basic use

Table of Contents

I. Setting up your computer to do the job

Testing that the migrator works on our computers

To test whether our computers are ready to execute the different steps of the migrators we can use an already-done migrator. Copy the tempvertnetprocessing.zip file that contains the already done migrator (you can ask VertNet folks for one to run the test). Unzip and copy the *tempvertnetprocessing* folder in *C:/*. It is important that this folder be in *C:*, as the migrator is set to interact with the files there though that route.

1. Migrator test part I: general compatibility and compatibility with Access

Open the folder "*tempvertnetprocessing*" in *C:/* and double click on the file "*1a - RunMigrators.bat.*" As part of the execution, a command window and Access will open, it will take some time to process (we can see queries running in Access, bottom right corner). Finally, if all is OK, Access will close automatically.

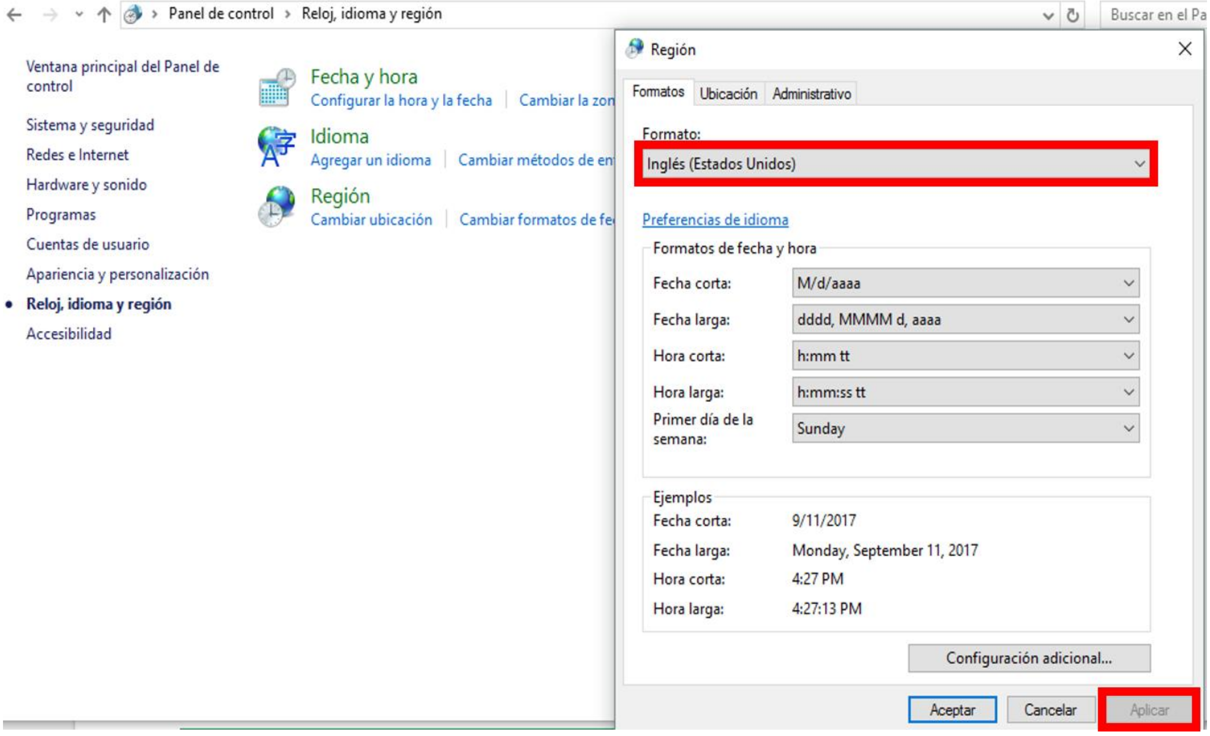
Possible errors

When we run a migrator a series of errors can occur. When this happens, a window will open showing the type of error encountered and the migrator will stop automatically. In these cases, we must close Access and the command window, solve the issue and execute the bat file again.

a. Macros disabled. If there are errors in the access permissions to run the macros, we have to: in Access go to Home→ Options → Trust Center → Trust Center settings → Macros setting → Enable all macros.
Close Access and execute the .bat file again.

b. Differences in regional settings. The migrator works well under an English-United States regional setting, but has issues under other regional settings in some other languages. This issue arises as some of the exports within the migrator use specifications that are in conflict with local settings that have "," as decimal indicator (instead of "."). To solve this issue, we must change the computer's regional setting. This change must be made from the control panel (not from the language icon on the tasks bar, that is not enough).

As shown in the following figure, choose "English (United States)" (or the equivalent written in your language, e.g., "Inglés (Estados Unidos)") and click on "Apply". Close and execute the .bat again.



Panel de control > Reloj, idioma y región

Región

Formatos | Ubicación | Administrativo

Formato: Inglés (Estados Unidos)

Preferencias de idioma

Formatos de fecha y hora

Fecha corta: M/d/aaaa

Fecha larga: dddd, MMMM d, aaaa

Hora corta: h:mm tt

Hora larga: h:mm:ss tt

Primer día de la semana: Sunday

Ejemplos

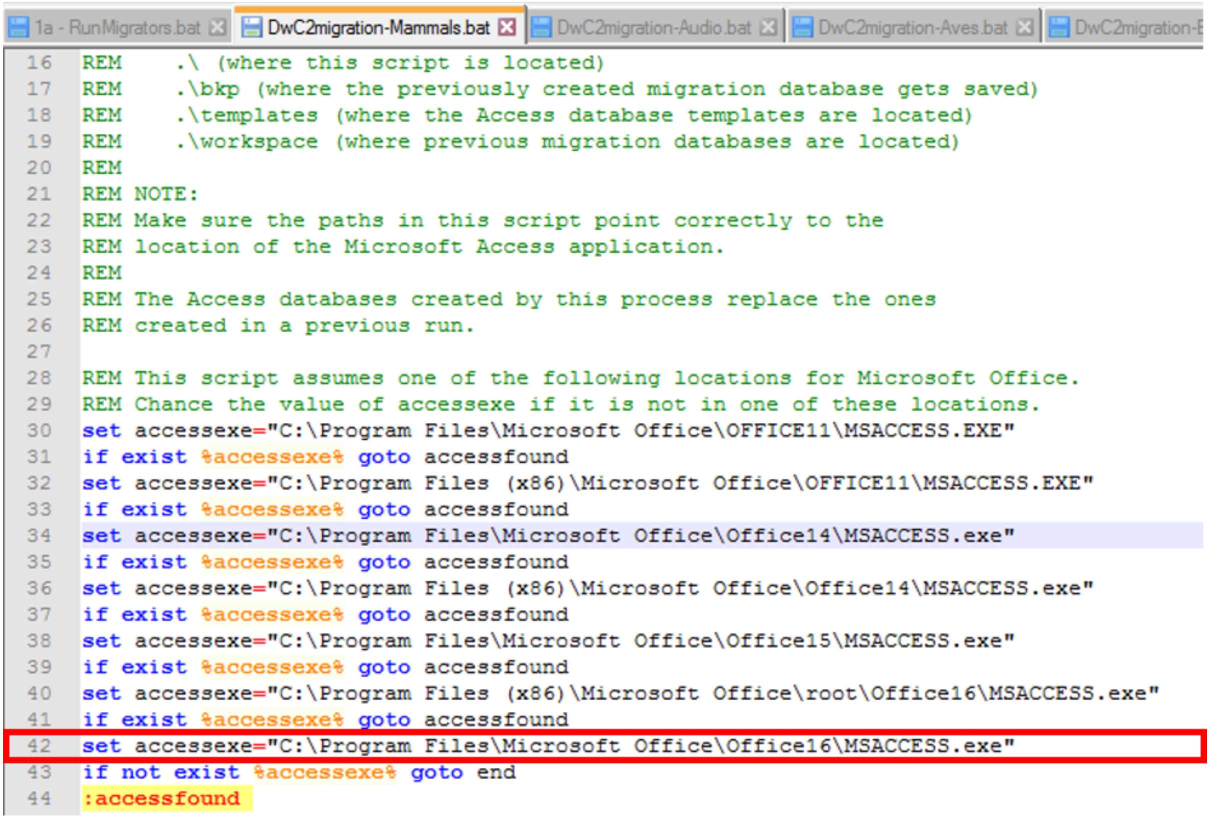
Fecha corta: 9/11/2017

Fecha larga: Monday, September 11, 2017

Hora corta: 4:27 PM

Hora larga: 4:27:13 PM

c. Access resides in a directory that is different from the one specified in the migrator. If when double-clicking the 1a - RunMigrators.bat. file the command window opens but Access does not open, it is possible that the directory in 1a - RunMigrators.bat that points to the Access.exe is incorrect. That is, the migrator is looking for Access in a place where it cannot be found. To solve this issue, we must change the access route to Access in all files in the migrator that use such route: all *DwC2migration-XXXX.bat* files (where XXXX is the migrator name, for example: *DwC2migration-Mammals.bat*, *DwC2migration-Aves.bat*, etc.) and in the "*1c - RunAggregators.bat*". We must change the route in ALL these files. To do so, we open each one in a text processor (DO NOT double click!, but do right click → open with…). There we'll see that there are several possible access routes to *MSACCESS.exe*. We replace one of these routes (it does not matter which one) with the access route that corresponds in our computer, as shown in the following figure, line 42.

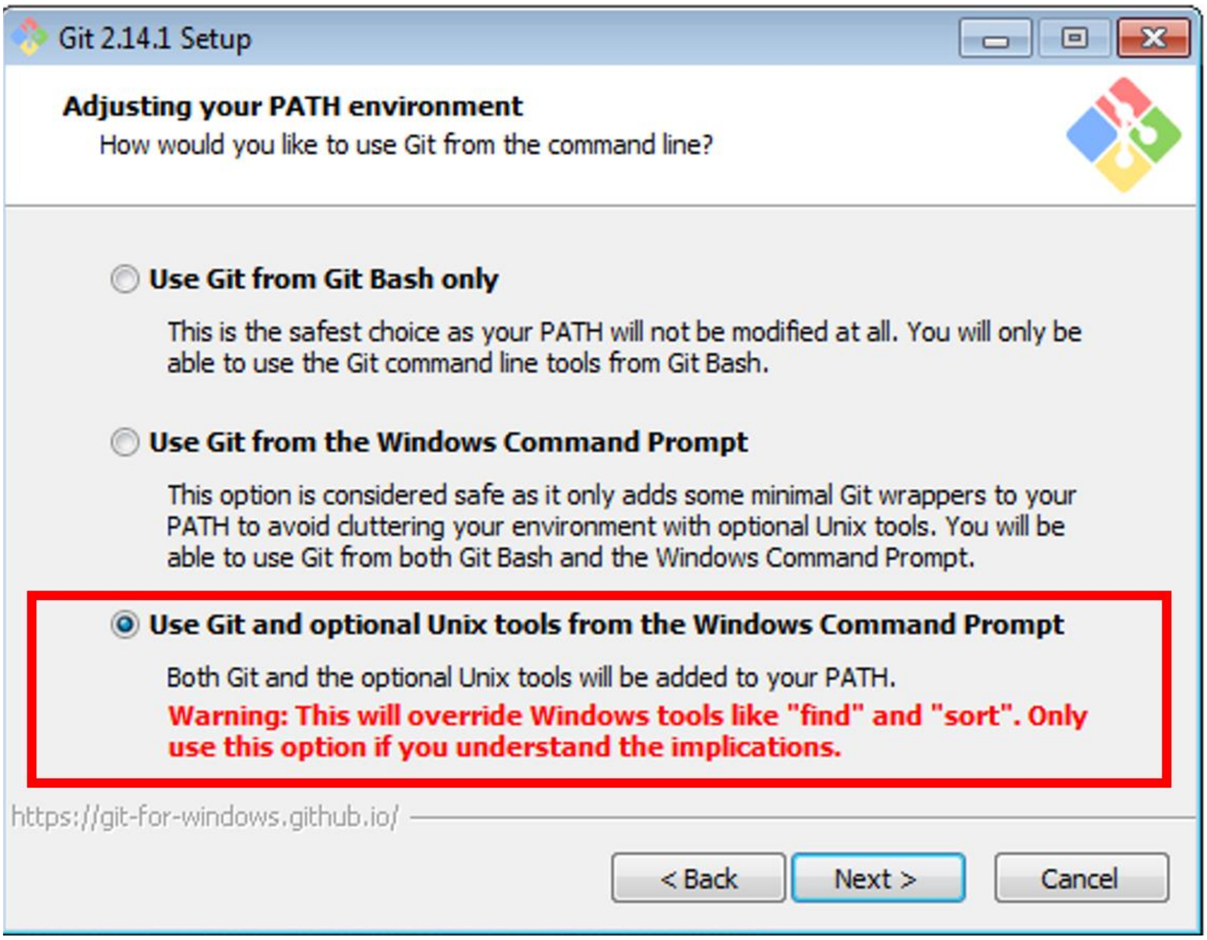


```
REM    .\  (where this script is located)
REM    .\bkp (where the previously created migration database gets saved)
REM    .\templates (where the Access database templates are located)
REM    .\workspace (where previous migration databases are located)
REM
REM NOTE:
REM Make sure the paths in this script point correctly to the
REM location of the Microsoft Access application.
REM
REM The Access databases created by this process replace the ones
REM created in a previous run.

REM This script assumes one of the following locations for Microsoft Office.
REM Chance the value of accessexe if it is not in one of these locations.
set accessexe="C:\Program Files\Microsoft Office\OFFICE11\MSACCESS.EXE"
if exist %accessexe% goto accessfound
set accessexe="C:\Program Files (x86)\Microsoft Office\OFFICE11\MSACCESS.EXE"
if exist %accessexe% goto accessfound
set accessexe="C:\Program Files\Microsoft Office\Office14\MSACCESS.exe"
if exist %accessexe% goto accessfound
set accessexe="C:\Program Files (x86)\Microsoft Office\Office14\MSACCESS.exe"
if exist %accessexe% goto accessfound
set accessexe="C:\Program Files\Microsoft Office\Office15\MSACCESS.exe"
if exist %accessexe% goto accessfound
set accessexe="C:\Program Files (x86)\Microsoft Office\root\Office16\MSACCESS.exe"
if exist %accessexe% goto accessfound
set accessexe="C:\Program Files\Microsoft Office\Office16\MSACCESS.exe"
if not exist %accessexe% goto end
:accessfound
```

2. Migrator test part II: Unix Utilities

The second part of the migrators ("*1c - RunAggregators.bat*") performs a series of processes that are not possible using only Access, and which require using Unix utilities, such as sh, sed, awk, etc. Therefore, we must have the utilities installed to be able to run the migrators without problems.

We will use the utilities that are included in git. Go the the git webpage and download the version compatible with your own operating system (check 32 or 64 bits): https://git-scm.com/downloads
Run the .exe file you downloaded to install git. As part of the installation, you will see a serie of options. Choose all the default options except from the following one:



Adjusting your PATH environment

How would you like to use Git from the command line?

○ **Use Git from Git Bash only**

This is the safest choice as your PATH will not be modified at all. You will only be able to use the Git command line tools from Git Bash.

○ **Use Git from the Windows Command Prompt**

This option is considered safe as it only adds some minimal Git wrappers to your PATH to avoid cluttering your environment with optional Unix tools. You will be able to use Git from both Git Bash and the Windows Command Prompt.

● **Use Git and optional Unix tools from the Windows Command Prompt**

Both Git and the optional Unix tools will be added to your PATH.

Warning: This will override Windows tools like "find" and "sort". Only use this option if you understand the implications.

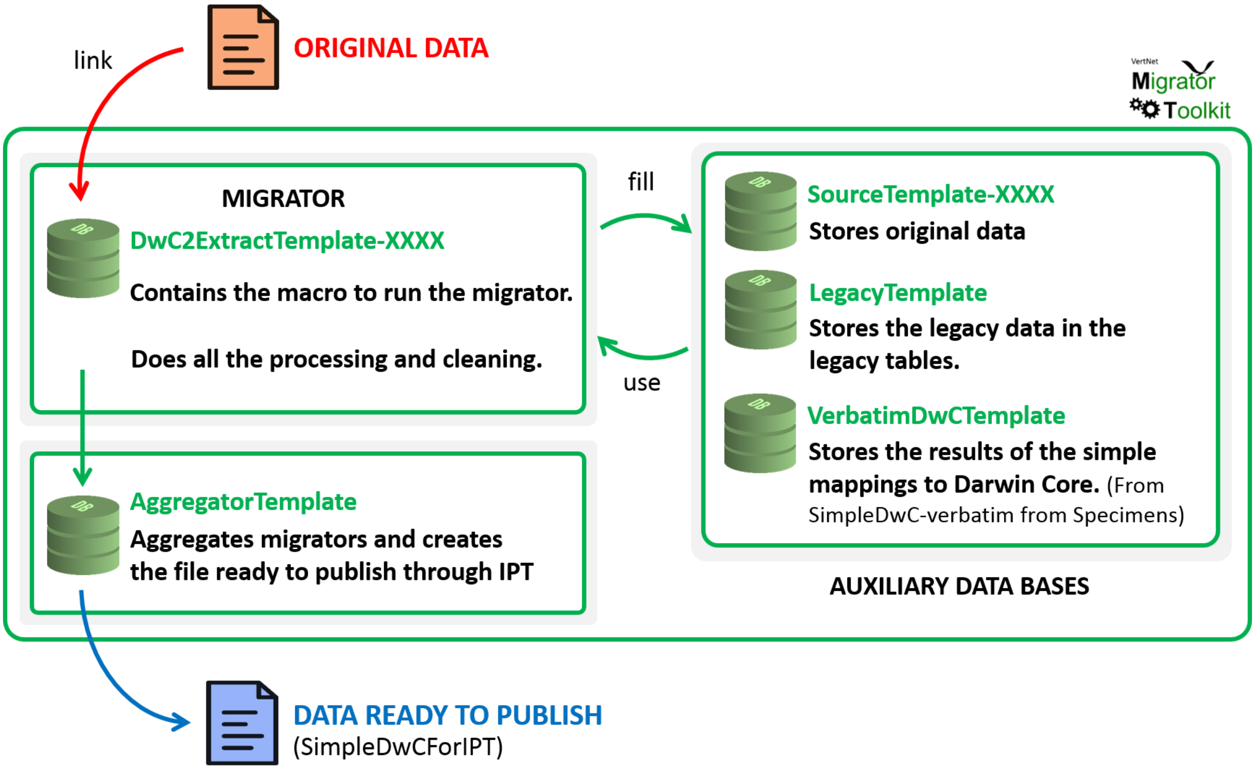
We must choose the last option so that we also install the Unix tools. Although there is an alert about find and sorting functions being affected, it will not alter the way in which we normally see our daily work folders; it will be using git instead of the default Windows function to perform these actions.

Once we finish git installation, we execute the second part of the migrator: “*1c - RunAggregators.bat*”. In doing so, if all is OK, the following will happen: a command window and Access will open automatically, they will process the migrator, and then they will both close automatically. To be sure that the process was successful, we go to the */ForIPT* folder in */tempvertnetprocessing*, and there we’ll see a series of files that were generated by the migrator.

II. Using the Migrators toolkit

A. Review of the migrator structure

The VertNet Migrator Toolkit uses a series of Access databases to process the data: DwC2ExtractTemplate-XXXX.mdb, SourceTemplate-XXXX.mdb, LegacyTemplate.mdb, VerbatimDwCTemplate.mdb and AggregatorTemplate.mdb. A brief explanation of what each of those does and how they are related is shown in the image below.


link
ORIGINAL DATA
VertNet Migrator Toolkit
MIGRATOR
DwC2ExtractTemplate-XXXX
Contains the macro to run the migrator.
Does all the processing and cleaning.
fill
use
SourceTemplate-XXXX
Stores original data
LegacyTemplate
Stores the legacy data in the legacy tables.
VerbatimDwCTemplate
Stores the results of the simple mappings to Darwin Core. (From SimpleDwC-verbatim from Specimens)
AggregatorTemplate
Aggregates migrators and creates the file ready to publish through IPT
AUXILIARY DATA BASES
DATA READY TO PUBLISH
(SimpleDwCForIPT)

The DwC2ExtractTemplate-XXXX.mdb is the main database in the migrator, and is the one where all the processing takes place. Within that database there are a series of tables, queries and macros. The details about those elements can be consulted here: Explanation - Migrator's Tables, Queries & Macros.

IMPORTANT: Below are listed a series of steps (**1 to 45**) that you should follow to run a basic migrator process. These steps should work just fine if you are migrating data from a single file (e.g., a single Excel spreadsheet). If the data from the source comes in several distinct files, you will require additional steps.

B. Preliminary steps, setting the migrator and linking the files

1. Get the latest migrator toolkit from GitHub at https://github.com/VertNet/toolkit, you can do it in two ways:

A. Click the green "Clone or download" button --> Download zip. Extract to an empty folder. The name of the folder will be toolkit. Rename the root folder to "tempvertnetprocessing" and move it to the C: drive. It must be in C: to match how the migrator is configured, or the migrator won't run.
B. If you are using git, clone the GitHub repository to your local machine. Any time you want to get the latest version from the master repository, use the command "git pull origin master" while in the directory where the repository was cloned.

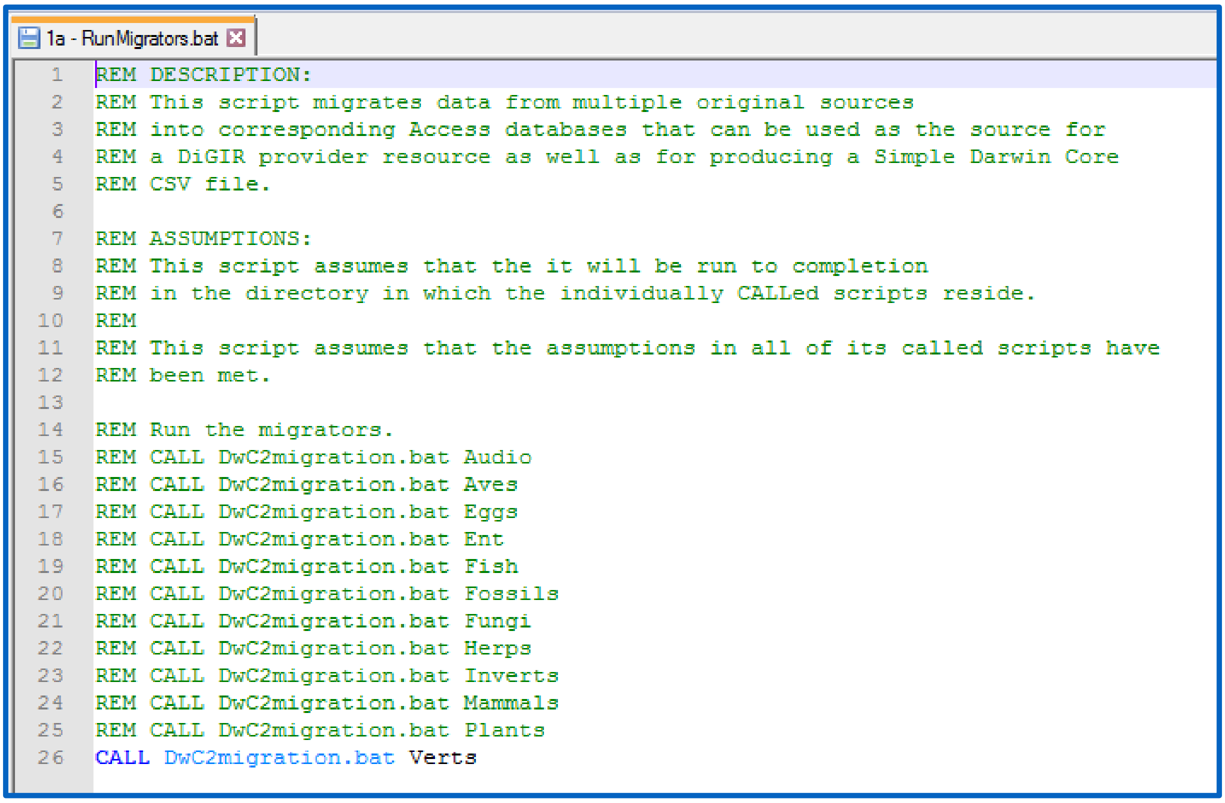
2. Copy the dataset you want to migrate into the "source" folder.

3. To get the latest vocabularies from the DwCVocabs GitHub repository, you can do it in two ways:

A. Go to https://github.com/tucotuco/DwCVocabs/tree/master/master, download and extract (unzip) the latest Vocabulary files (2: VocabulariesMaster.zip and VocabulariesManager.zip) into the "source" folder, so that the migrator can write data to the files.
B. If you are using git, clone the GitHub repository to your local machine. Any time you want to get the latest version from the master repository, use the command "git pull origin master" while in the DwCVocabs directory.

Keep in mind that if you are keeping a separate vocabularies file that has not been merged to the master repository in GitHub, that master will not have your latest changes.

4. Set the correct migrators to be run for the collection in the script "1a - RunMigrators.bat". To do so, DO NOT double click it, instead, open it in a text editor (right click --> open with). Set the appropriate migrator(s) to "CALL" (e.g., "CALL DwC2Migration.bat Herps"), and leave the rest commented out with "REM" so that they do not run. Save and close the bat file.

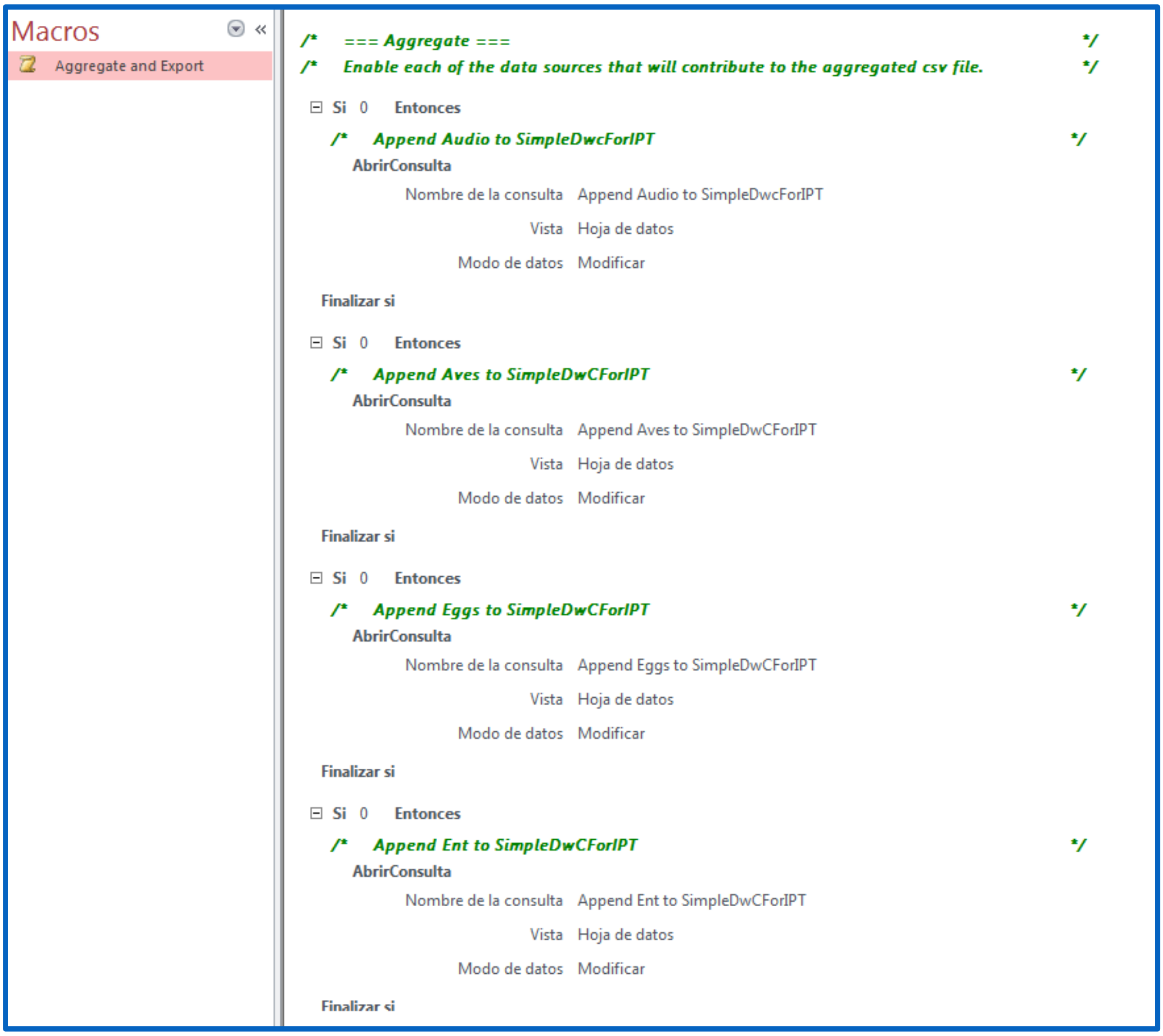


```
REM DESCRIPTION:
REM This script migrates data from multiple original sources
REM into corresponding Access databases that can be used as the source for
REM a DiGIR provider resource as well as for producing a Simple Darwin Core
REM CSV file.

REM ASSUMPTIONS:
REM This script assumes that the it will be run to completion
REM in the directory in which the individually CALLed scripts reside.
REM
REM This script assumes that the assumptions in all of its called scripts have
REM been met.

REM Run the migrators.
REM CALL DwC2migration.bat Audio
REM CALL DwC2migration.bat Aves
REM CALL DwC2migration.bat Eggs
REM CALL DwC2migration.bat Ent
REM CALL DwC2migration.bat Fish
REM CALL DwC2migration.bat Fossils
REM CALL DwC2migration.bat Fungi
REM CALL DwC2migration.bat Herps
REM CALL DwC2migration.bat Inverts
REM CALL DwC2migration.bat Mammals
REM CALL DwC2migration.bat Plants
CALL DwC2migration.bat Verts
```

5. In the “templates” folder, edit AggregatorTemplate.mdb to run the macro to assemble the data from the appropriate migrator(s) and save the results as a csv file.

Open AggregatorTemplate.mdb (in the “templates” folder) in Access and edit the macro “Aggregate and Export”. To do so, DO NOT double click, as that will run the macro, do right click --> design view.

- Turn on the append queries to add data from the results of the desired migrators (e.g., “Append Herps to SimpleDwCForIPT”) by changing the condition from 0 to 1. Turn off the Verts migrator, which is the default query enabled, unless you want to use it.
- Save the macro changes and then close the database.



```
/*    === Aggregate ===                                                                    */
/*    Enable each of the data sources that will contribute to the aggregated csv file.     */
⊟ Si  0     Entonces
   /*    Append Audio to SimpleDwcForIPT                                                   */
      AbrirConsulta
            Nombre de la consulta   Append Audio to SimpleDwcForIPT
                            Vista   Hoja de datos
                    Modo de datos   Modificar
 Finalizar si
⊟ Si  0     Entonces
   /*    Append Aves to SimpleDwCForIPT                                                    */
      AbrirConsulta
            Nombre de la consulta   Append Aves to SimpleDwCForIPT
                            Vista   Hoja de datos
                    Modo de datos   Modificar
 Finalizar si
⊟ Si  0     Entonces
   /*    Append Eggs to SimpleDwCForIPT                                                    */
      AbrirConsulta
            Nombre de la consulta   Append Eggs to SimpleDwCForIPT
                            Vista   Hoja de datos
                    Modo de datos   Modificar
 Finalizar si
⊟ Si  0     Entonces
   /*    Append Ent to SimpleDwCForIPT                                                     */
      AbrirConsulta
            Nombre de la consulta   Append Ent to SimpleDwCForIPT
                            Vista   Hoja de datos
                    Modo de datos   Modificar
 Finalizar si
```

6. Open SourceTemplate-XXXX.mdb in Access (the one in the “templates” folder), make a linked table from the source data and call it "Occurrence". To do so, follow these steps:

1. In the menu External Data, import data (from text, from Excel, from Access, etc.). If you are importing from Excel, it is recommended that you first export the data from Excel as csv. In the window that will open select “Link to the data source””.

2. In Browse look for the source file (the original data that you copied into the ”source” folder in step **2.**). Click OK.

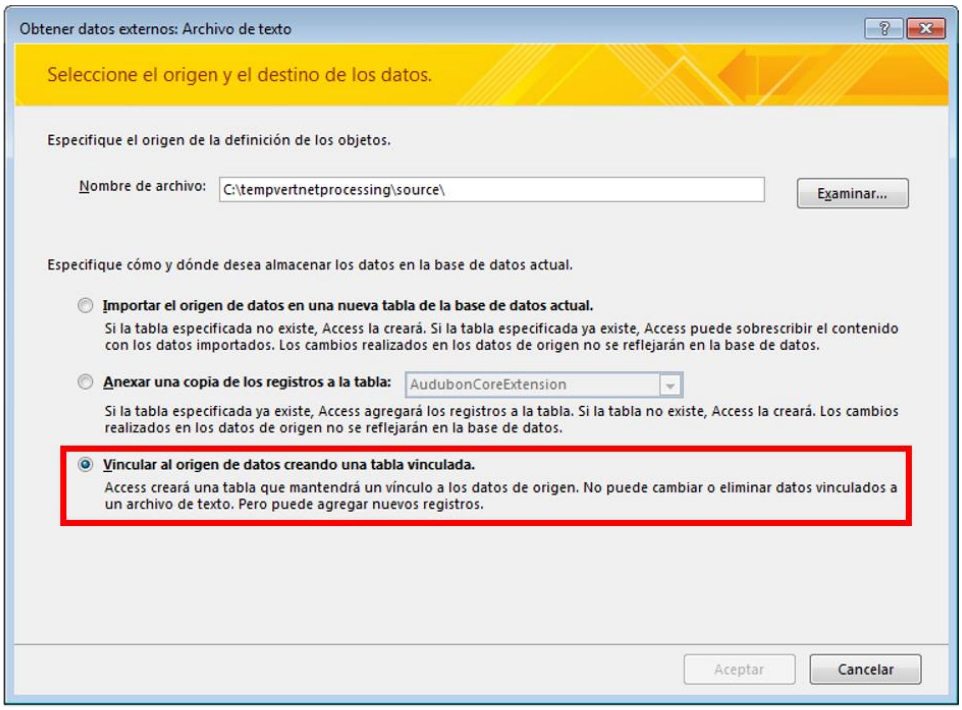
3. In the following windows, choose “Delimited” for the width, and “First Row Contains Column Headers”. You may also need to select a text qualifier (“ or ‘).

4. If the file is .TXT go to Advanced (in any of the panes) and set Code Page to the correct encoding. In general, the encoding is UTF-8, but could be another one. Please CHECK the encoding of your dataset, for example by looking at the few records that the importing window shows, and further check after importation by examining the data. Cases in which errors can appear are accents and special characters, typical of certain languages (e.g., ES, FR).

5. Set Data Types for all fields in the table to “text” (short text), or “memo” (long text) if the field might contain a lot of characters (i.e., more than 255; e.g., comments fields, such as georeferenceRemarks or locality should be long text).

5. Remember to call the linked table “Occurrence”. Finish.

-NOTE: In some files, CSVs usually, text may have “” around it, and Access may say that it doesn’t like some of the headers, this resulting in added/unnecessary columns. If this is true, you need to change the text qualifier. Usually, you can see if there are “” in the data just by reviewing the file in a text editor. Review the table Occurrence for diacriticals and other special characters to see if they are rendered correctly. If not, either the encoding chosen for the file is incorrect, or the data have gone through a history of imports and exports that did not respect the correct encoding.



Obtener datos externos: Archivo de texto

Seleccione el origen y el destino de los datos.

Especifique el origen de la definición de los objetos.

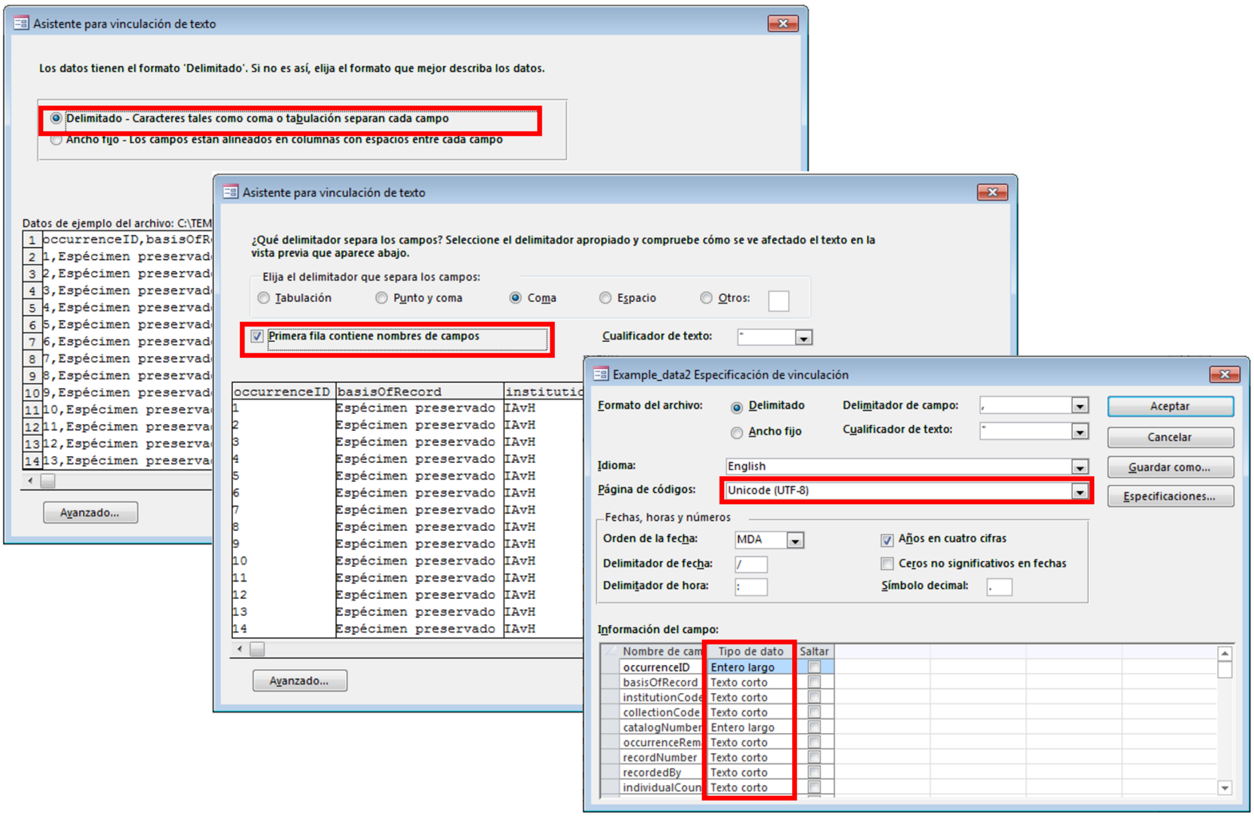
Nombre de archivo: C:\tempvertnetprocessing\source\

Especifique cómo y dónde desea almacenar los datos en la base de datos actual.

○ **Importar el origen de datos en una nueva tabla de la base de datos actual.**
Si la tabla especificada no existe, Access la creará. Si la tabla especificada ya existe, Access puede sobrescribir el contenido con los datos importados. Los cambios realizados en los datos de origen no se reflejarán en la base de datos.

○ **Anexar una copia de los registros a la tabla:** AudubonCoreExtension
Si la tabla especificada ya existe, Access agregará los registros a la tabla. Si la tabla no existe, Access la creará. Los cambios realizados en los datos de origen no se reflejarán en la base de datos.

● **Vincular al origen de datos creando una tabla vinculada.**
Access creará una tabla que mantendrá un vínculo a los datos de origen. No puede cambiar o eliminar datos vinculados a un archivo de texto. Pero puede agregar nuevos registros.



Asistente para vinculación de texto

Los datos tienen el formato 'Delimitado'. Si no es así, elija el formato que mejor describa los datos.

● Delimitado - Caracteres tales como coma o tabulación separan cada campo

○ Ancho fijo - Los campos estan alineados en columnas con espacios entre cada campo

Datos de ejemplo del archivo: C:\TEM

```
1  occurrenceID,basisOfR
2  1,Espécimen preservad
3  2,Espécimen preservad
4  3,Espécimen preservad
5  4,Espécimen preservad
6  5,Espécimen preservad
7  6,Espécimen preservad
8  7,Espécimen preservad
9  8,Espécimen preservad
10 9,Espécimen preservad
11 10,Espécimen preserva
12 11,Espécimen preserva
13 12,Espécimen preserva
14 13,Espécimen preserva
```

Asistente para vinculación de texto

¿Qué delimitador separa los campos? Seleccione el delimitador apropiado y compruebe cómo se ve afectado el texto en la vista previa que aparece abajo.

Elija el delimitador que separa los campos:

○ Tabulación ○ Punto y coma ● Coma ○ Espacio ○ Otros: □

☑ Primera fila contiene nombres de campos

Cualificador de texto: "

```
occurrenceID basisOfRecord        institutic
1            Espécimen preservado IAvH
2            Espécimen preservado IAvH
3            Espécimen preservado IAvH
4            Espécimen preservado IAvH
5            Espécimen preservado IAvH
6            Espécimen preservado IAvH
7            Espécimen preservado IAvH
8            Espécimen preservado IAvH
9            Espécimen preservado IAvH
10           Espécimen preservado IAvH
11           Espécimen preservado IAvH
12           Espécimen preservado IAvH
13           Espécimen preservado IAvH
14           Espécimen preservado IAvH
```

Example_data2 Especificación de vinculación

Formato del archivo: ● Delimitado ○ Ancho fijo

Delimitador de campo: ,

Cualificador de texto: "

Idioma: English

Página de códigos: Unicode (UTF-8)

Fechas, horas y números

Orden de la fecha: MDA

Delimitador de fecha: /

Delimitador de hora: :

☑ Años en cuatro cifras

□ Ceros no significativos en fechas

Símbolo decimal: .

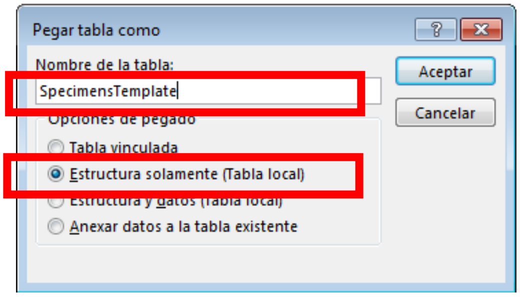
Información del campo:

Nombre de cam	Tipo de dato	Saltar
occurrenceID	Entero largo	□
basisOfRecord	Texto corto	□
institutionCode	Texto corto	□
collectionCode	Texto corto	□
catalogNumber	Entero largo	□
occurrenceRem	Texto corto	□
recordNumber	Texto corto	□
recordedBy	Texto corto	□
individualCoun	Texto corto	□

7. Remove the SpecimensTemplate table. (Right click --> Delete)

SpecimensTemplate

8. Copy the structure (only) of the Occurrence table to a new SpecimensTemplate table. To do so, select the Occurrence table, right click and select “Copy”, then right click again and select “Paste”, rename as “SpecimensTemplate” and choose “Structure Only”.



Pegar tabla como

Nombre de la tabla: SpecimensTemplate

Opciones de pegado

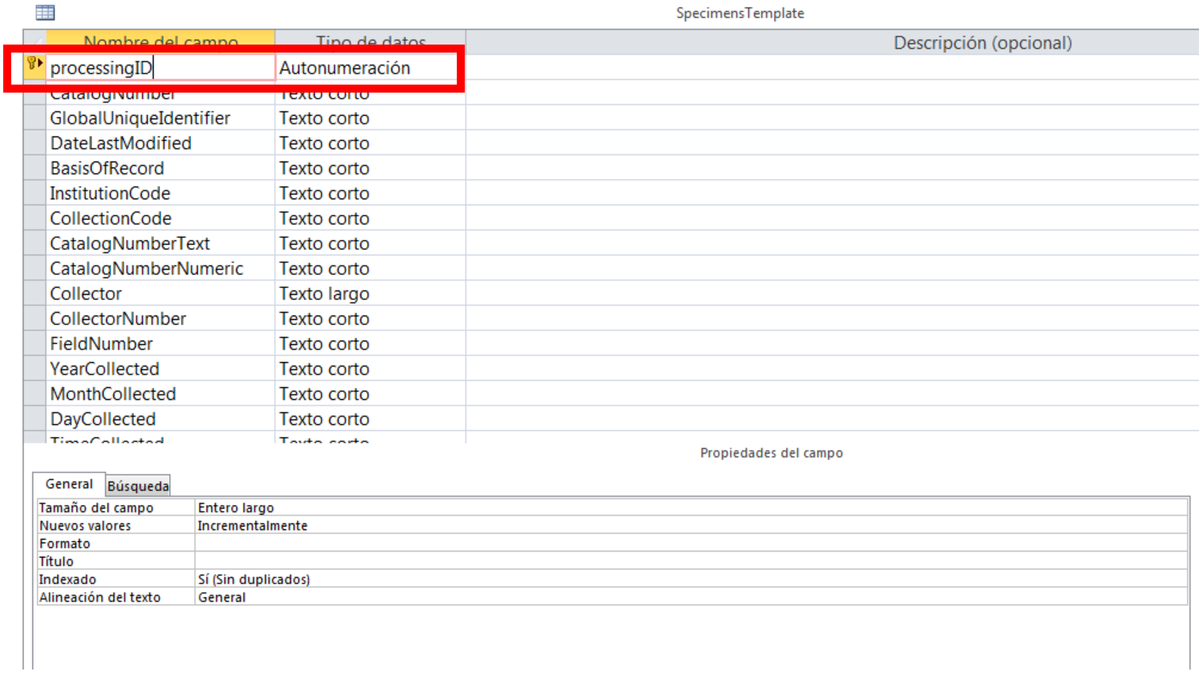
○ Tabla vinculada

● Estructura solamente (Tabla local)

○ Estructura y datos (Tabla local)

○ Anexar datos a la tabla existente

9. Edit the SpecimensTemplate table you just created: open the table in design view (right click --> design view) Add a row to the top of the table, name it “processingID” and then give it a data type ‘AutoNumber’. Now make that new row a Primary Key: right-click on the grey square by the name of the field, select Primary Key. Now the row will have the primary key icon (a little key). Save and close.



SpecimensTemplate

Nombre del campo	Tipo de datos	Descripción (opcional)
processingID	Autonumeración	
CatalogNumber	Texto corto	
GlobalUniqueIdentifier	Texto corto	
DateLastModified	Texto corto	
BasisOfRecord	Texto corto	
InstitutionCode	Texto corto	
CollectionCode	Texto corto	
CatalogNumberText	Texto corto	
CatalogNumberNumeric	Texto corto	
Collector	Texto largo	
CollectorNumber	Texto corto	
FieldNumber	Texto corto	
YearCollected	Texto corto	
MonthCollected	Texto corto	
DayCollected	Texto corto	
TimeCollected	Texto corto	

Propiedades del campo

Tamaño del campo	Entero largo
Nuevos valores	Incrementalmente
Formato	
Título	
Indexado	Sí (Sin duplicados)
Alineación del texto	General

10. Remove the Specimens table (right click --> delete).

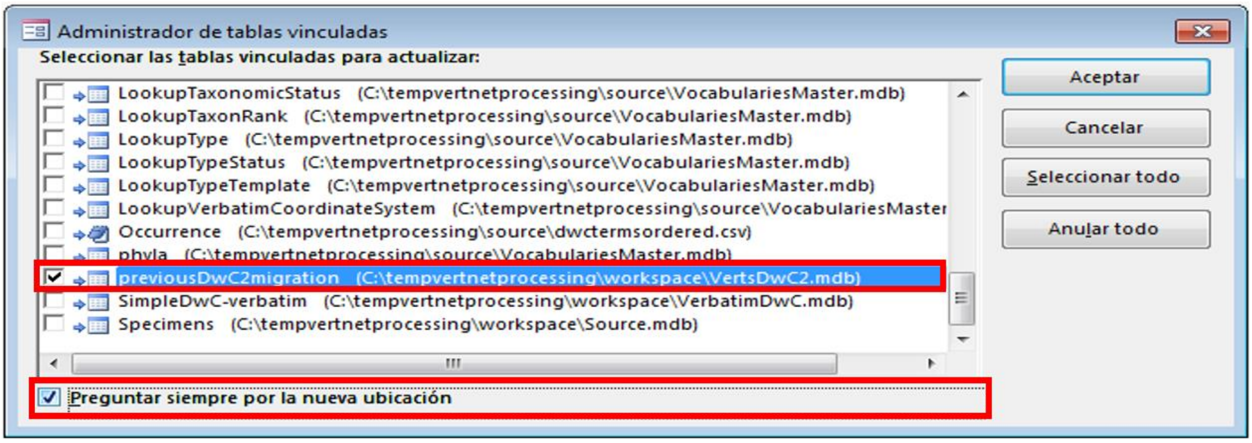
11. Copy the SpecimensTemplate table to a table called Specimens (structure only, as you did in step **8** with the Occurrence Table, though since the SpecimensTemplate table is empty you could also choose structure and data and it would be the same at this point).

12. Close the SourceTemplate-XXXX.mdb data base. Copy this template data base (all the data base) from the “templates” folder and paste it into the “workspace” folder. Rename it “Source-XXXX.mdb”.

13. In the “templates” folder, rename the template database DwC2ExtractTemplate-XXXX.mdb to match the scope of the data set for which the migrator will be made, where XXXX is one of “Audio”, "Aves", "Eggs", “Ent”, "Fish", "Fossils", “Fungi”, "Herps", “Inverts”, "Mammals", “Plants”, "Verts", etc. If you use any other template name, you will have to modify the script "1a - RunMigrators.bat" to add a new CALL to invoke it., create a script to run that migrator using the pattern “Dwc2migration-XXXX.bat”, add queries for the reports in the macro "Do DwC2 Migration" and then modify the database AggregatorTemplate.mdb to be able to process the results.

14. Open DwC2ExtractTemplate-XXXX.mdb in Access (the one in the “templates” folder, which was renamed in step **13.**). There you can organize the left toolbar to see the different types of objects (tables, queries, forms, macros, etc.) and to order them by different criteria (type of object, date created, etc). You can select All Access Objects to display them all. You can right click one of the section bars, e.g., Tables, to sort by Name.

15. Link table previousDwC2migration to the correct migrator (in the “workspace” folder). To do so, go to External Data, click Linked Table Manager, scroll down to the table previousDwC2migration and select the option “Always prompt for new location”. Click OK. A folder will open for you to choose a new file to link the table to. Link your table to the appropriate database in the folder "workspace" (e.g., HerpsDwC2.mdb). Then close the Linked Table Manager window.



Administrador de tablas vinculadas

Seleccionar las tablas vinculadas para actualizar:

- □ LookupTaxonomicStatus (C:\tempvertnetprocessing\source\VocabulariesMaster.mdb)
- □ LookupTaxonRank (C:\tempvertnetprocessing\source\VocabulariesMaster.mdb)
- □ LookupType (C:\tempvertnetprocessing\source\VocabulariesMaster.mdb)
- □ LookupTypeStatus (C:\tempvertnetprocessing\source\VocabulariesMaster.mdb)
- □ LookupTypeTemplate (C:\tempvertnetprocessing\source\VocabulariesMaster.mdb)
- □ LookupVerbatimCoordinateSystem (C:\tempvertnetprocessing\source\VocabulariesMaster
- □ Occurrence (C:\tempvertnetprocessing\source\dwctermsordered.csv)
- □ phyla (C:\tempvertnetprocessing\source\VocabulariesMaster.mdb)
- ☑ previousDwC2migration (C:\tempvertnetprocessing\workspace\VertsDwC2.mdb)
- □ SimpleDwC-verbatim (C:\tempvertnetprocessing\workspace\VerbatimDwC.mdb)
- □ Specimens (C:\tempvertnetprocessing\workspace\Source.mdb)

☑ Preguntar siempre por la nueva ubicación

16. In this template database you are working on, remove the linked table called "Occurrence" (right click --> Delete), the one that looks like this:

Occurrence

17. Repeat step **6.** (the one you performed on the SourceTemplate-XXXX.mdb) on this data base (DwC2ExtractTemplate-XXXX). That is, make a linked table from the source data and call it "Occurrence".

18. Edit the contents of the LookupCollectionMetadata table to match the collection being migrated. To do so, double click in the table and make the respective changes in the elements of the table according to the collection information (metadata).
Confirm that the data set does not have a mix of fossil, preserved, and observation data, since basisOfRecord might need to be mapped in another way if so. If all the records have the same basisOfRecord, you can capture it in this LookupCollectionMetadata table. If they don’t do not map it here, and leave it for the mapping step further on.

Open LookupCollectionMetadata table

LookupCollectionMetadata

- Check GBIF for institutionCode
- Check GRBio (http://grbio.org/) for institutionID - use CoolURI if present, if not use institutional LSID
- Check GBIF for collectionCode
- Add the appropriate Creative Commons license link to the license field: CC0, CC BY, CC BY-NC.
- Add VN Norms link to the accessRights field - http://www.vertnet.org/resources/norms.html

- Add the collectionID from GRBio if listed.

19. If you want to test the migrator with a small number of records (recommended until the migrator can run completely without errors), use a filter on the Source to include only a small number. If the dataset is too large to run through the migrator (>~1M records), then it may be necessary to make two or more migrators, each of which uses a source filter that accepts a subset of the original data. Open the query "Source" in design view and edit it to filter the table "Occurrence", if necessary. This means: in the query “Source”, you’ll see that the whole table “Occurrence” is linked; if you wanted just some of the fields included, as opposed to all of them, you could filter them here. Filters can be useful for two things:

FILTER 1> on fields. For example, you could exclude fields that you know providers would not be interested in sharing (e.g., fields that capture the location of a specimen within a collection – “drawer number”). This does not necessarily need to happen here, you can wait until the mapping step and not map those fields.

FILTER 2> on records. For example, if you had a dataset that is too large to process, you may want to cut it in parts and apply the process to the chunks separately. One way to do this is to identify a field that can easily cut the dataset in chunks without leaving any record out (e.g., a numeric id field, which is populated for every record: you could use ># as a first filter, operate, then do the same for the complementary filter <=#).

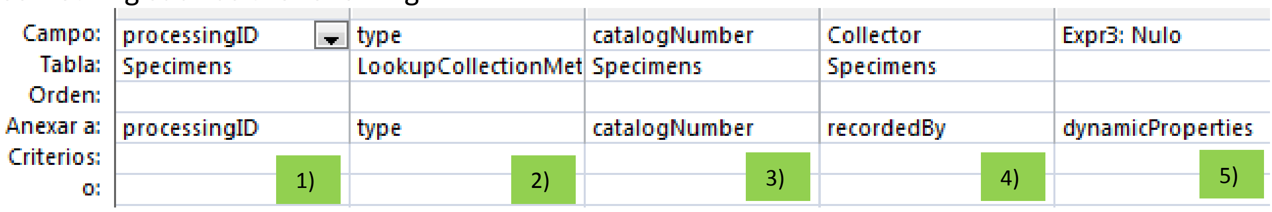
NOTE: If you are working with a big data set, it is recommended that you do set a filter temporarily, so that you run the process on a subset until you make sure the migrator works. This will save you a lot of time.

C. Now it’s time to map and customize the migrator, Iuhuuu!!

20. Mapping

Open the query called "SimpleDwC-verbatim from Specimens" in design view (do not run it now). This will be the template to turn Specimens into the first pass of verbatim data in SimpleDwC, that is: the macro will run on this query (step **28**).

In the design view of the “SimpleDwC-verbatim from Specimens" query you will see something such as the following:



Campo:	processingID	type	catalogNumber	Collector	Expr3: Nulo
Tabla:	Specimens	LookupCollectionMet	Specimens	Specimens	
Orden:					
Anexar a:	processingID	type	catalogNumber	recordedBy	dynamicProperties
Criterios:					
o:					

First line: field in the occurrence or in the LookupCollectionMetadata table, the ones you are going to map.
Second line: the table from which each of the fields in the first row is coming from (Occurrence or LookupCollectionMetadata table). This will be blank if the first line contains a constant value or an expression rather than just a field name.
Fourth line: the fields into which you are going to map.

Some examples:

1) processingID. This field is not a DwC field! But, it is needed for the migrator to connect tables during some of the steps. Do not touch it. Leave the mapping as default.
2) Fields that map directly to a field in the LookupCollectionMetadata table.
3) Fields that map directly to a field in the Specimens table.
4) Fields have a non-DwC name, but that can be mapped directly (because they are equivalent).
5) Fields that do not map directly. For example, in the table above there are no dynamic properties in the Specimens table, nor any other field with a different name that can be exclusively mapped to it. In these cases, leave the field null (Expr#: Null) or just delete that field (right click, “cut”) from the design view. Do not delete the field from the Specimens table.

The mapping should include only fields that can be mapped directly. Any other processing should occur later in the process. For example, if there is no eventDate but there are beginDate and endDate fields, do not map them, there is a specific function to do that later in the process. Check that the query is visible without errors by trying to switch to Datasheet view, without running it (as that would attempt to append data to the Specimens table; that is: do not double click it!). If the query changes to Datasheet view without errors, that is what we want at this point. Note that the Datasheet view will not show any records. That is OK.

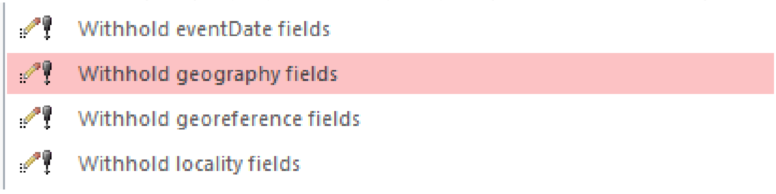
Remove all columns that do not apply to the data (e.g., geological context for non-fossils). This is just to keep this table clean and simple (all DwC terms will appear anyway in the end result).

21. “Hiding” data: encumbrances

In the macro "Do DwC2 Migration", edit and enable encumbrances if necessary. This is meant to hide (withhold) certain fields or records that the provider might not want to share. There are encumbrances for: records, geography fields, locality fields, and georeferencing fields. Other encumbrances can be created as well.

- If you do not want any fields or records to be hidden, set the condition in the “Do DwC2 Migration” macro for “Withhold fields or records” to 0.

- If you want some fields to be hidden, then you need to edit the corresponding encumbrance query (each query can have conditions) and set the condition in the “Withhold fields or records” macro for that encumbrance to 1. For example: open the “Withhold geography fields” query (right click --> Design view) and modify it as necessary.



Withhold eventDate fields
Withhold geography fields
Withhold georeference fields
Withhold locality fields

22. Prepending all verbatim geography into verbatimLocality

This step is to keep all geographic verbatim information “safe” in verbatimLocality. Remember that this will happen if later (step **28**) in the macro "Do DwC2 Migration", you enable the corresponding queries called "Prepend GGGG to verbatimLocality", where GGGG corresponds with a higher geography field, to make verbatimLocality preserve all original higher geography whether in Darwin Core fields or not. (see step **28**). This is necessary because the migrator will do vocabulary lookups to standardize the Darwin Core geography fields.

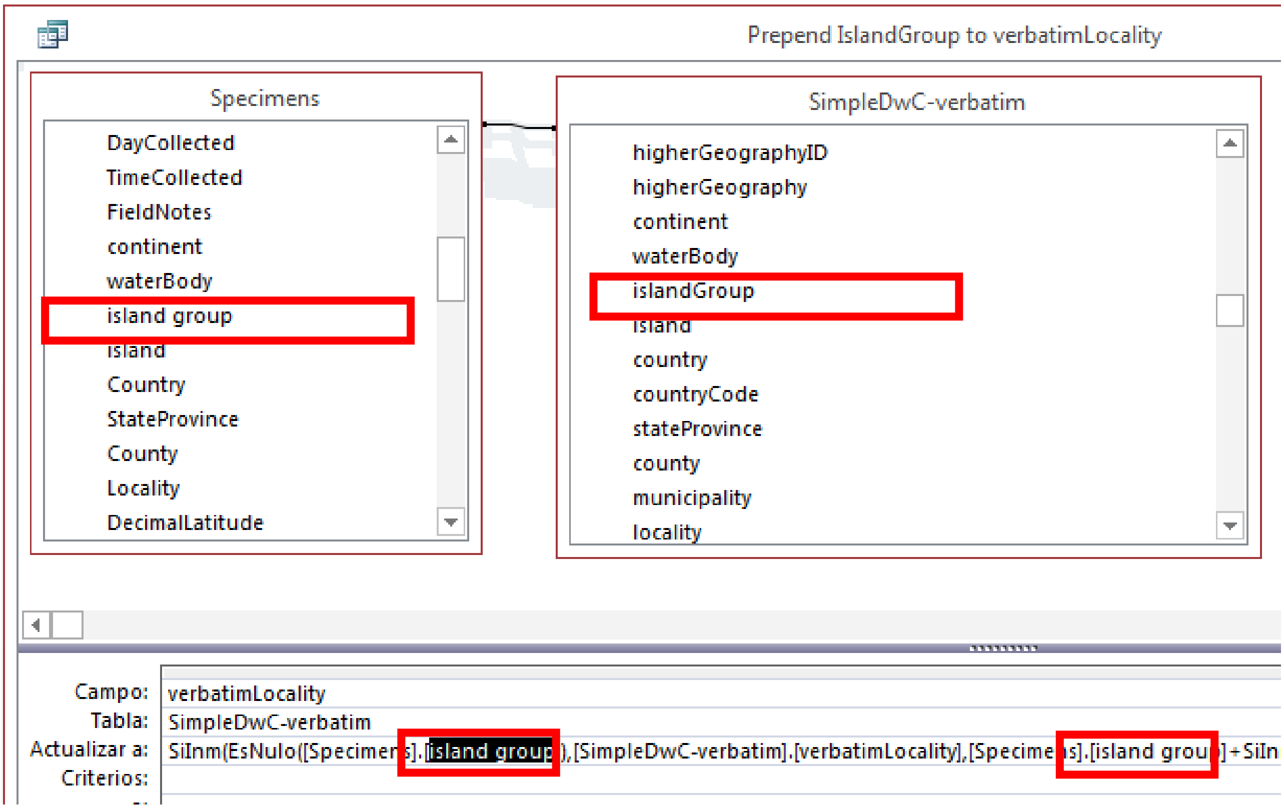
For example, “Prepend Waterbody to verbatimLocality” will put the waterbody data into the verbatimLocality field, concatenated with any other field you are also prepending. Since this is going to make a concatenation, it is important to have the fields ordered correctly, from least to most specific in the verbatimLocality field, and therefore in reverse order of queries because they are prepending to the verbatimLocality.

NOTE: the fields you are prepending are those in the original data. So, be careful to use the names of the fields that appear in those original data (Specimens table), and not the names you would give after mapping. Example: suppose the original data has a field “island group” (separated by white space), which maps to dwc:islandGroup. For the prepend, use “island group”, and not “islandGroup”: open the corresponding "Prepend islandGroup to verbatimLocality" query using Design View and change the “Update To” field to reflect “island group” and not "islandGroup”. If the field name is the same as the Darwin Core field name, you will not have to make any changes to the query, as the field you want is already the one in the Specimens table (e.g., [Specimens].[islandGroup].

The default expression looks like the following:

```
SiInm(EsNulo([Specimens].[island group]),[SimpleDwC-verbatim].[verbatimLocality],[Specimens].[island group]+SiInm(EsNulo([SimpleDwC-verbatim].[verbatimLocality]),""," | "+[SimpleDwC-verbatim].[verbatimLocality]))
```

Remember to change the term every time it appears in the expression.



Prepend IslandGroup to verbatimLocality

Specimens	SimpleDwC-verbatim
DayCollected	higherGeographyID
TimeCollected	higherGeography
FieldNotes	continent
continent	waterBody
waterBody	islandGroup
island group	island
island	country
Country	countryCode
StateProvince	stateProvince
County	county
Locality	municipality
DecimalLatitude	locality

Campo:	verbatimLocality
Tabla:	SimpleDwC-verbatim
Actualizar a:	SiInm(EsNulo([Specimens].[island group]),[SimpleDwC-verbatim].[verbatimLocality],[Specimens].[island group]+SiIn
Criterios:	

Repeat this process for every geographic field, checking the names of the fields in the Specimens table.

Add a new query for any higher geographic field that is not already represented in the prepend list (e.g., drainage, national park, etc.). Remember that if any new queries are added, such as one for Drainage, it needs to be added to the Macro (see below, step **28**) before the macro is run. To add a query, copy/paste/rename one that you already have (on the left panel, right click on any prepend query --> copy, right click --> paste). Edit the new query in design view to match the corresponding field (e.g., Drainage).

23. Putting all verbatim taxonomy into higherClassification

Edit the query called "Update higherClassification with verbatim taxonomy" to make higherClassification preserve all original classification. To do so, open the query in design view. In the "Update To" field, the default expression looks like the following:

Recortar(SiInm(EsNulo(specimens.kingdom),"",specimens.kingdom)+" | "+SiInm(EsNulo(specimens.phylum),"",specimens.phylum)+" | "+SiInm(EsNulo(specimens.class),"",specimens.class)+" | "+SiInm(EsNulo(specimens.order),"",specimens.order)+" | "+SiInm(EsNulo(specimens.family),"",specimens.family)+" | "+SiInm(EsNulo(specimens.genus),"",specimens.genus))

In that expression, note that the names of the fields are coming from the Specimens table, and therefore they should have the actual names that the Specimens table uses for each one (e.g., if the original field is “Phyla” instead of “Phylum”, you should call it “Phyla”, and that part of the expression will be: +SiInm(EsNulo(specimens.phyla),"",specimens.phyla)+

Take out from that expression all fields that do not apply for the migrator you are doing.

Add classifications that are not represented in the query, such as suborder or subfamily. Separate each rank with " | " and go from least specific (kingdom) to most specific (subgenus).

Keep all fields that correspond to a taxon rank down to subgenus (i.e., do not include specificEpithet nor infraspecificEpithet, nor var, etc.). Do not include scientificName (or its equivalent).
Close the query and SAVE it.

Dealing with fields that do not map directly to DwC

24. Edit the queries called "LegacyXXX from Specimens" (where XXXX is one of "Coordinates", "CoordinateUncertainty", "Dates", "Depths", "Elevations", or "ScientificNames") (these are all append queries with a green+ in their icon) to populate Legacy Data tables with fields that don't fit directly into SimpleDwC-verbatim or for fields that require special processing (multi-step or joins to other tables) to transform the original data into the recommended Darwin Core fields and formats. If any Legacy queries are edited, the Macro may need to be updated to turn them on (see step **28**).

25. Add and customize queries for dynamicProperties if necessary. There is a series of “Update dynamicProperties for XXXX” queries already available, where XXXX is the property/characteristic (e.g. measurements, plumage, etc). If present, update relevant queries to reflect the data set field names and match to appropriate DwC fields in each query.
If any dynamicProperties are used, be sure to enable them in the macro “Dynamic Properties” and to set the condition to run “Dynamic Properties” to 1 in the “Do DwC2 migration” macro (see step **28**).

26. Edit the query "Update occurrence - preparations for tissues" to make preparations include tissue information if it does not already. Make sure the names of the fields match those in the original data in the “Update To” field in the query.

27. Open the query "Report - missing catalogNumber" in design view and edit it to use the catalog number field from the source.

Now it's time to turn ON and OFF the migrator's functions!!! :D

28. Open the macro "Do DwC2 Migration" in design mode (right click-->design view). Do not run it yet! (that is, do not double click on it). Turn on and off steps as appropriate and edit any queries that require customization.
While turning ON/OFF, you may have to interact with tables and queries. Most of the changes you have already done in previous steps, but you may check as you turn ON/OFF each step.

In the macro you will see a series of steps, each one with a brief description of what it does, organized in sections. Each step can be turned ON or OFF by changing the condition from 0 to 1 and vice versa. It can be useful to keep the original data open somewhere to understand what it actually contains and how each step would affect it. If the original data does not have some particular columns (e.g., date), then all the actions pertaining to such columns should be turned off (e.g., all actions on dates), otherwise there will be an error because the macro is trying to do something on data that does not exist.

Some steps can have one or both of the following marks in their comments:

!!!	signifies that the step is required and SHOULD NOT BE TURNED OFF.
$$$	signifies that the query may require customization and should be checked for compatibility with the source data structure and contents.

In some cases, both sets of symbols are present, which means that the step is required and needs customization. Any step that does not have $$$ in its comment should not have to be customized. Any step that has only !!! should not be edited, nor turned off.

Some steps to check/amend for sure are listed below:

- Encumbrances Section = macro "Withhold records and fields"
  - Macro should be off unless there are to be encumbrances. Queries in macro should be on and customized only if there are to be encumbrances.
- Prepend Section = "Preprend…"
  - Turn ON all fields that are used. If Waterbody, for example, is present in the data set, the query "Prepend waterBody to verbatimLocality" must be turned ON.
  - Turn OFF all fields not used in the data set, those that do not have a correspondence in the Specimens table. For example, if the Specimens table does not have any field corresponding to municipality (a third-level geopolitical subdivision within country, with that name or any other), set the condition for the query "Prepend municipality to verbatimLocality" to 0. If you don't, the macro with stop with an error when that step is executed.
  - Add a new action in the macro for any added queries in the process. Example: if a drainage field exists in the original data, a query "Prepend Drainage to verbatimLocality" should have been made and needs to be added to the macro. In this case, since this is going to make a concatenation, it is important to have the fields ordered correctly, try to go from least specific (continent) to most specific (in the macro you'll see them listed the other way

around, the most specific on top). To add an action to the macro: copy and paste a macro entry of the corresponding section. Edit the new macro entry to invoke that new query.

- Update Section = "Update Simple DwC - …" Generally no changes are needed to these actions, but check!
- Legacy Sections. Enable or disable each paired query “LegacyXXX from Specimens” and associated macro “Legacy XXX” from the following list:
    - Coordinates - Turn ON if data set include any geographic coordinates fields.
    - CoordinateUncertainties - Turn ON if data set includes any coordinate uncertainty fields.
    - Dates - Turn ON if data set includes any event date fields (year, month, day, eventDate, verbatimEventDate or equivalents).
    - Depths - turn ON if Depths fields are in use.
    - Elevations - turn ON if Elevation fields are in use.
    - ScientificNames - turn ON if a equivalent to scientificName is in use and there are no separate fields for genus and lower ranks.
- DynamicProperties Section - turn ON the macro “Dynamic Properties” if any dynamic properties from the original data will be captured. Customize and enable any such dynamicProrpties in the macro “Dynamic Properties”.
- Clean Up SimpleDwC Table Sections - generic processing, usually just leave as is, but check any that have $$$ in the comment.
- Audubon Media Extension Section - turn ON the macro “Audubon Media Extension”if media links are present in the data.
    - If there are media data to be processed from associatedMedia field into the Audubon Media Extension, turn ON the appropriate steps in the macro “Audubon Media Extension”. If a separate media file is provided, additional customization will need to be done to link in those data and populate the table "AudubonCoreExtension", including modifying the query "Append media to AudubonCoreExtension" based on the linked file.
- RelationshipResource Section - One common use of the ResourceRelationship Extension is to notify GBIF that the global unique identifiers for records have changed. For example, if there is a change in how the occurrenceIDs are assigned from an older published version of a data set, the ResourceRelationshipExtension will have to be populated with the old and new identifiers. Turn ON the macro “Resource Relationship Extension” if there are relationships to add.
    - If there are resource relationships to be added, turn ON the appropriate steps in the macro “Resource Relationship Extension”under "Add occurrenceIDs to RelationshipResourceExtension" in the macro.

D. Run the migrator! For the… first time!!!

29. Test the migrator by running the script "1a) RunMigrators.bat" , in the "tempvertnetprocessing" folder with the appropriate CALL(s) enabled.
If the test fails, Click Stop/Cancel. Important: note in the pop up box which query was responsible for the error and what the cause was. The most common cause for errors is a reference to a field name that either does not exist or is ambiguous because it exists in more than one referenced table. The query that caused the error will have to be amended in the template database (not the temp database that is open while running the .bat script). Close the database and wait for the script to finish.

E. Yeah… I know… errors… it stopped in the middle of the process, didn't it…

30. Fix the error that occurred in the previous step by editing the template database in the "templates" folder (not the database made in the "workspace" folder by the "1a) RunMigrators" script).
To do so, open templates\(specific Template) and look for the query that threw the error in step 27 (e.g., if the error suggested that the specimens.eventDate was not found while executing the query "LegacyDates from Specimens", open that query to check and update the mapping).

- Correct mappings for fields present in the data set. Close query and save.
- Once corrections are made, attempt to test again at the start of Step **29.**

31. Iterate through steps **29.** and **30.** until all errors have been fixed.

F. Checking everything turned out fine

32. Check the results of the migration by reviewing the data in the SimpleDwC table in the migrated database in the folder "workspace". Confirm all data have been migrated to the correct fields with the correct transformations.

33. Make any adjustments needed to populate the SimpleDwC table as completely as possible by editing queries and turning them on or off in the macro "Do DwC2 migration" and any other necessary macros that it invokes.

34. Run the script "1) RunMigrators.bat" with the appropriate CALL within turned on (e.g., "CALL DwC2migration.bat Aves") one more time.

35. Iterate through steps **32.** through **33.** until all data populates SimpleDwC correctly.

G. Invoking multiple migrators

36. Sometimes it is necessary to create distinct migrators for data from distinct sources within an institution and to aggregate those datasets in one data set. All of these can be run together using the “1a - RunMigrators.bat” file after turning on and off appropriate migrators as described in Step **4**.

37. Run 1c - RunAggregators.bat.
NOTE: Do this step even if you don’t do Step **36.** It is needed to get back reports on blank line issues and to generate the CSV file(s) that are ready to publish to Darwin Core Archives.

- When complete, open the ForIPT folder to find the new txt files reporting on blank lines, encoding and other issues.
- If, when reviewing these txt files, any records are reported, they should be reported to the source so that they can be corrected. These records will have been fixed in the migration, but should be fixed at the source.
- Next check the Reports folder. There are two reports that require special attention from the data provider:
    - Review Reports-LinesWithNonprintingCharacters - report any that appear to the source.
    - Review Reports-duplicate catalogNumber - reporting any to the source if the source uses catalog numbers.

  NOTE: opening these reports directly with Excel will not show diacriticals or hidden characters correctly, so they should be reviewed using Notepad++ or other editor that respects encoding and allows non-printing characters to be viewed, if necessary.

H. Checking and resolving vocabularies

38. Check Vocabularies. To do so, open the database “source\VocabulariesManager.mdb”. Select view to see all macros. Run the macro "Unchecked Local Vocabs", which will display windows for each vocabulary that is normally tracked and resolved.

39. Edit the open vocabularies that show values that need to be checked to map to standard versions of the terms. Some of these may be deferred, such as preparations and lifeStage, which are incredibly diverse. These vocabulary views are not shown by default by the “Unchecked Local Vocabs” macro. They can be turned on by setting the condition for them to 1 in the marco “Unchecked Local Vocabs”.

I. Re-running the migrator to include the newly resolved vocabs

40. Run the script "1a - RunMigrators.bat" again to make a copy of the SimpleDwC table with vocabulary lookups included.

J. Put together the datasets

41. Run the script "1b - RunAggregators.bat" after migrators for all distinct collections have been finished up to this step to make a final aggregated migrated data set.

42. Compare the counts on output to SimpleDwCForVertNet.csv with the expected number from the processed collection SimpleDwC tables. Resolve any discrepancies.

K. Final steps: reports and data to publish

43. Archive the error report files (called "Report -*" in the reports folder) to send to the data publisher. Alert the publisher about these reports so they can improve their data at the source, if desired.

44. Zip and then copy the SimpleDwCForIPT.csv file to send for review by the data publisher. Alert them that opening the file directly in Excel will likely misinterpret the encoding, which is UTF-8. Provide guidance on how to preserve the encoding when opening in Excel (see http://vertnet.org/resources/downloadsinexcelguide.html for guidance). When ready to zip reports for review, you do not need to include “processed.csv” nor “LineWithNonprintingCharacters”, but do include an appropriate copy of the Report Explanation document.

45. Upload the file SimpleDwCForIPT.zip to the resource on the IPT when it is authorized to do so by the data publisher.

CONGRATULATIONS!!! :D :D