

# Herramienta de Migración de datos de VertNet's Instrucciones de uso básico

# **Tabla de Contenidos**

I. Configurando la computadora para hacer el trabajo	2
1. Prueba de migrador parte I: compatibilidad general y con Access	2
2. Prueba de migrador parte II: Utilidades Unix	
II. Usando la herramienta Migrador	<del>(</del>
A. Revisión de la estructura del migrador	6
B. Pasos preliminares, configurando el migrador y vinculando los archivos	7
C. Ahora es tiempo de mapear y personalizar el migrador, yuhuu!!	15
D. Ejecute el migrador! Por primera vez!!!	21
E. Sí lo sé errores paró en la mitad del proceso, ¿no fue así?	21
F. Comprobando que todo salió bien	22
G. Invocando múltiples migradores	22
H. Revisando y resolviendo vocabularios	23
I. Ejecutando nuevamente el migrador para incluir los nuevos vocabularios resueltos	23
J. Ensamblando los conjuntos de datos	23
K. Pasos Finales: reportes y datos para publicación	24
FELICITACIONES III · D · D	24



# I. Configurando la computadora para hacer el trabajo

# Probando que el migrador funciona en nuestras computadoras

Para probar que nuestras computadoras están preparadas para ejecutar los distintos pasos del migrador, utilizamos un migrador ya hecho. Copiamos el archivo tempvertnetprocessing.zip que contiene el migrador (Paula lo paso en una usb). Se descomprime y se copia en *C:/*. Es importante que la carpeta *tempvertnetprocessing* esté en el disco *C:*, porque el migrador la interpela desde esa ruta de acceso.

#### 1. Prueba de migrador parte I: compatibilidad general y con Access

Abrir la carpeta "tempvertnetprocessing" en *C:/* y dar doble click al archivo "1a - RunMigrators.bat." Como parte de la ejecución, se abrirán una ventana de comandos y Access y demora un tiempo en procesar (podemos ver en Access, en la esquina inferior derecha, que está ejecutando consultas). Finalmente, si todo salió, Access se cierra solo.

#### **Posibles errores**

Cuando se ejecuta el migrador pueden aparecer una serie de errores. Cuando esto sucede, aparecerá una ventana indicando el tipo de error y se suspenderá el migrador automáticamente. En estos casos, debemos cerrar Access y la ventana de comandos, solucionar el problema correspondiente y ejecutar nuevamente el archivo .bat.

a. Macros deshabilitados. Si hay error de permisos para los macros, se debe ir en Access a
 Inicio→ Opciones → Centro de confianza → Configuración del centro de confianza → Configuración de macros → Habilitar todos los macros.

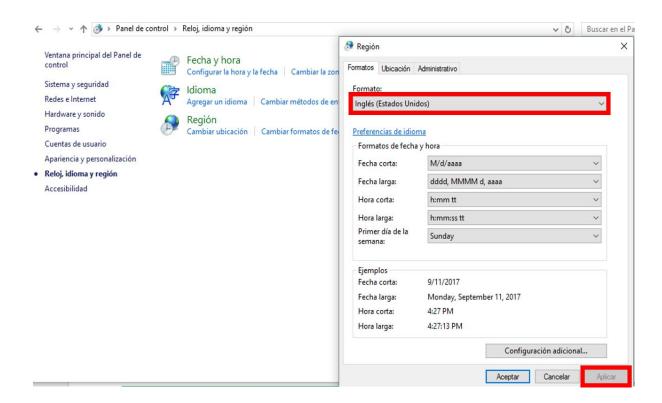
Cerrar Access y ejecutar nuevamente el archivo .bat.

**b. Diferencias en configuración regional.** El migrador funciona bien con configuración regional en inglés-Estados Unidos, pero no con configuraciones regionales en algunos otros idiomas. Esto es porque algunas exportaciones dentro del migrador utilizan especificaciones que están en conflicto con configuraciones locales que tienen "," como indicador decimal (en vez de "."). Para solucionar este problema, debemos cambiar la configuración regional de la computadora. Este cambio lo debemos hacer desde el panel de control (no desde el ícono de idioma de la barra de tareas, eso no alcanza).

En la siguiente figura, escoger "Inglés (Estados Unidos)" y hacer click en "Aplicar". Luego cerrar y ejecutar nuevamente el archivo .bat.







c. Access está en un directorio diferente al especificado en el migrador. Si al dar doble click al 1a - RunMigrators.bat. y correr la ventana de ejecución no se abre Access, el posible error es el directorio que apunta al ejecutable de Access en el archivo 1´a - RunMigrators.bat. Es decir, el migrador está buscando el programa Access en un lugar donde no está. Para solucionar este problema, hay que cambiar la ruta de acceso a Access de todos los archivos del migrador que lo usan: todos los archivos *DwC2migration-XXXX.bat* (donde XXXX es el nombre del migrador, por ejemplo: *DwC2migration-Mammals.bat*, *DwC2migration-Aves.bat*, etc.) y en el "1c - *RunMigrators.bat*". Hay que hacerlo en TODOS estos archivos. Para ello, abrimos cada uno en un procesador de texto (NO hacer doble click! sino click derecho → abrir con...). Allí veremos que hay varias rutas de acceso posibles a *MSACCESS.exe*. Reemplazamos una de esas rutas (no importa cual) por la ruta de acceso que corresponda en nuestra computadora, como se ve en la línea 42 de la siguiente imágen.





```
📑 1a - Run Migrators.bat 🔀 📙 DwC2migration-Mammals.bat 🔀 📙 DwC2migration-Audio.bat 🔀 📙 DwC2migration-Aves.bat 🔀 🛗 DwC2migration-Aves.bat
            .\ (where this script is located)
 16 REM
 17 REM
            .\bkp (where the previously created migration database gets saved)
 18 REM
          .\templates (where the Access database templates are located)
 19
    REM
            .\workspace (where previous migration databases are located)
 20
    REM
 21 REM NOTE:
 22 REM Make sure the paths in this script point correctly to the
    REM location of the Microsoft Access application.
 23
    REM
 25 REM The Access databases created by this process replace the ones
 26 REM created in a previous run.
 27
 28
    REM This script assumes one of the following locations for Microsoft Office.
 29 REM Chance the value of accessexe if it is not in one of these locations.
 30 set accessexe="C:\Program Files\Microsoft Office\OFFICE11\MSACCESS.EXE"
    if exist %accessexe% goto accessfound
 31
     set accessexe="C:\Program Files (x86)\Microsoft Office\OFFICE11\MSACCESS.EXE"
 33 if exist %accessexe% goto accessfound
 34 set accessexe="C:\Program Files\Microsoft Office\Office14\MSACCESS.exe"
 35
    if exist %accessexe% goto accessfound
     set accessexe="C:\Program Files (x86)\Microsoft Office14\MSACCESS.exe"
 37
     if exist %accessexe% goto accessfound
 38 set accessexe="C:\Program Files\Microsoft Office\Office15\MSACCESS.exe"
 39
    if exist %accessexe% goto accessfound
     set accessexe="C:\Program Files (x86)\Microsoft Office\root\Office16\MSACCESS.exe"
     if exist %accessexe% goto accessfound
42 set accessexe="C:\Program Files\Microsoft Office\Office16\MSACCESS.exe"
 43 if not exist %accessexe% goto end
44 :accessfound
```

#### 2. Prueba de migrador parte II: Utilidades Unix

La segunda etapa de los migradores ("1c - RunAggregators.bat") lleva a cabo una serie de procesos que no son posibles utilizando únicamente Access, y que requieren utilizar utilidades de Unix, como sh, sed, awk, etc. Por lo tanto, debemos tener instaladas las utilidades para que el migrador pueda ejecutarse sin problemas.

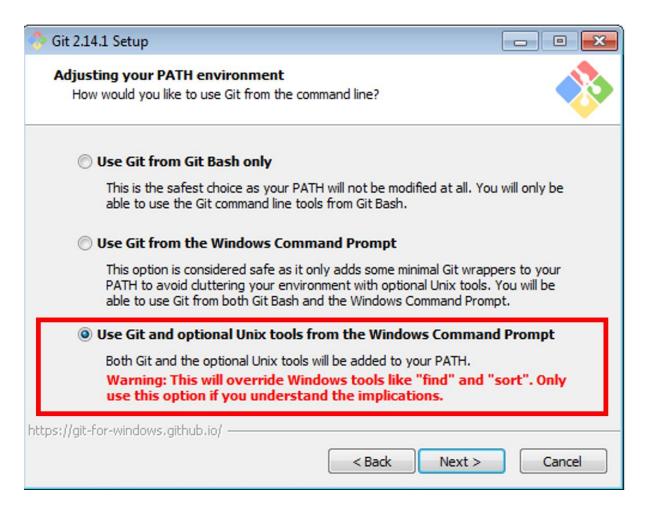
Usaremos las utilidades que vienen con git.

Ir a la página de git y descargar el que corresponda al sistema operativo propio (chequear 32 o 64 bits): https://git-scm.com/downloads

Ejecutar el archivo .exe descargado para instalar git. En la instalación aparecerán una serie de opciones. Escogeremos todas las marcadas por default salvo la siguiente:







Tenemos que elegir la última opción para que instale también las herramientas Unix. Si bien nos alerta que las funciones de búsqueda y orden pueden ser afectadas, no altera la manera en que lo vemos en las carpetas normales, utiliza la función de git en vez de la default de windows.

Una vez que terminamos la instalación de git, ejecutamos la segunda parte del migrador: "1c - RunAggregators.bat".

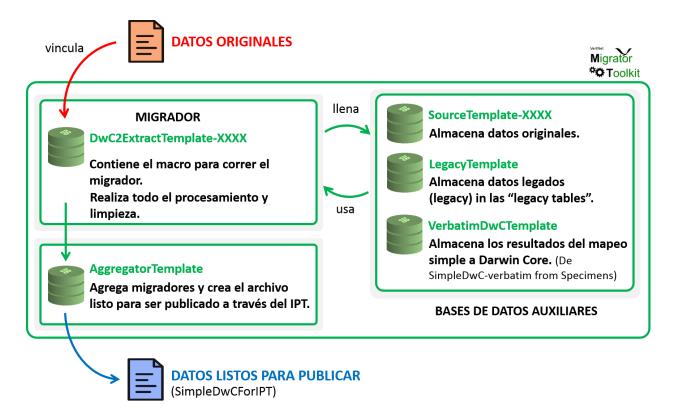
Al hacerlo, si todo sale bien ocurre lo siguiente: se abren una ventana de comando y Access automáticamente, procesan, y luego se cierran ambos automáticamente. Para estar seguros de que el proceso fue exitoso, vamos a la carpeta /ForIPT dentro de /tempvertnetprocessing, y allí veremos una serie de archivos que fueron generados por el migrador.



# II. Usando la herramienta Migrador

# A. Revisión de la estructura del migrador

La herramienta Migrador de VertNet utiliza una serie de bases de datos de Access para procesar los datos: DwC2ExtractTemplate-XXXX.mdb, SourceTemplate-XXXX.mdb, LegacyTemplate.mdb, VerbatimDwCTemplate.mdb y AggregatorTemplate.mdb. En la imagen a continuación se muestra una breve explicación sobre qué hace cada una de ellas y cómo se relacionan entre sí.



El DwC2ExtractTemplate-XXXX.mdb es la base de datos principal del migrador, y es aquella donde ocurre todo el procesamiento. Dentro de esta base de datos hay una serie de tablas, consultas y macros. Los detalles sobre estos elementos pueden ser consultados aquí: Explanation - Migrator's Tables, Queries & Macros.

**IMPORTANTE:** A continuación son listados una serie de pasos (1 a 45) que usted debería seguir para ejecutar un proceso de migrador básico. Estos pasos deberían funcionar correctamente si usted está migrando datos desde un archivo único (e.g., una única planilla Excel). Si los datos provienen de la fuente en varios archivos distintos, requerirá pasos adicionales.





# B. Pasos preliminares, configurando el migrador y vinculando los archivos

- **1.** Obtenga la última versión de la herramienta migrador en GitHub desde <a href="https://github.com/VertNet/toolkit">https://github.com/VertNet/toolkit</a>, puede hacerlo de dos maneras:
  - A. Haga "Click" en el botón verde "Clone or download" --> Download zip. Extraiga el contenido del archivo comprimido a una carpeta vacía. El nombre de la carpeta será "toolkit-develop". Renombre la carpeta raíz como "tempvertnetprocessing" y muévala al disco C:. Debe estar en C: para poder corresponderse con cómo el migrador está configurado, de otro modo el migrador no se ejecutará.
  - B. Si utiliza git, clone el repositorio en GitHub a su máquina local. Siempre que desee obtener la última versión desde el repositorio madre, utilice el comando "git pull origin develop" mientras esté en el directorio donde el repositorio fue clonado.
- 2. Copie el conjunto de datos que desea migrar a la carpeta "source".
- **3.** Para obtener los vocabularios más actualizados del repositorio DwCVocabs en GitHub, puede escoger una de estas dos opciones:
  - A. Vaya a <a href="https://github.com/tucotuco/DwCVocabs/tree/master/master">https://github.com/tucotuco/DwCVocabs/tree/master/master</a>, descargue y extraiga (descomprima) las últimas versiones de los archivos de Vocabularios (2: VocabulariesMaster.zip y VocabulariesManager.zip) en la carpeta "source", para que el migrador pueda escribir datos en estos archivos.
  - B. Si utiliza git, clone el repositorio en GitHub a su máquina local. Siempre que desee obtener la última versión desde el repositorio madre, use el comando "git pull origin master" mientras esté en el directorio DwCVocabs.

Tenga presente que si usted está manteniendo archivos de vocabularios separados que no han sido integrados al repositorio madre en GitHub, dichos archivos madre no contendrán los cambios más actuales que usted haya realizado.

**4.** Configure los migradores correctos a ejecutar para la colección en el script "1a - RunMigrators.bat". Para ello, NO haga doble click en él, en cambio, ábralo en un editor de texto (click derecho--> abrir con). Configure el/los migrador/es apropiados a "CALL" (e.g., "CALL DwC2Migration.bat Herps"), y deje el resto comentado con "REM" para que no se ejecuten. Guarde los cambios y cierre el archivo bat.





```
🔚 1a - RunMigrators.bat 🔣
     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:
  8 REM This script assumes that the it will be run to completion
     REM in the directory in which the individually CALLed scripts reside.
 11
    REM This script assumes that the assumptions in all of its called scripts have
 12 REM been met.
 14 REM Run the migrators.
    REM CALL DwC2migration.bat Audio
 16 REM CALL DwC2migration.bat Aves
 17 REM CALL DwC2migration.bat Eggs
 18 REM CALL DwC2migration.bat Ent
 19 REM CALL DwC2migration.bat Fish
 20 REM CALL DwC2migration.bat Fossils
     REM CALL DwC2migration.bat Fungi
 22 REM CALL DwC2migration.bat Herps
 23 REM CALL DwC2migration.bat Inverts
 24 REM CALL DwC2migration.bat Mammals
 25 REM CALL DwC2migration.bat Plants
 26 CALL DwC2migration.bat Verts
```

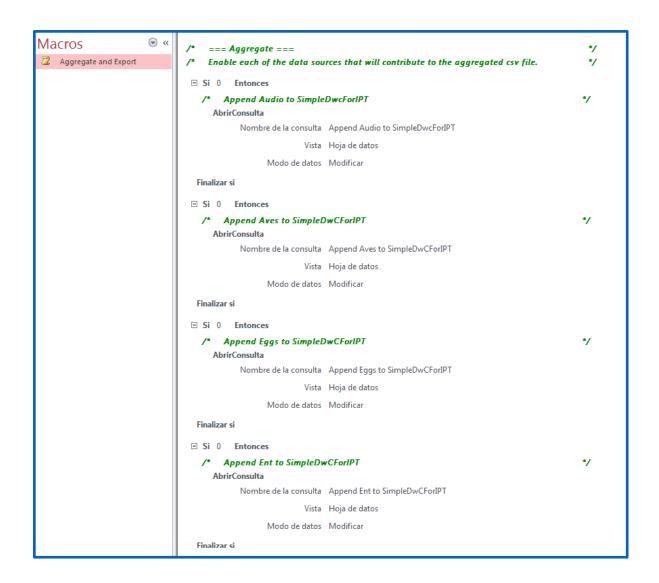
5. En la carpeta "templates", edite el archivo AggregatorTemplate.mdb para ejecutar el macro que ensamble los datos desde el/los migrador/es apropiados y guarda los resultados como un archivo csv.

Abra AggregatorTemplate.mdb (en la carpeta "templates") en Access y edite el macro "Aggregate and Export". Para ello, NO haga doble click sobre él, dado que ello ejecutará el macro, haga click derecho --> vista diseño.

- Prenda las consultas de datos anexados ("append queries") para agregar datos desde los resultados de los migradores deseados (e.g., "Append Herps to SimpleDwCForIPT") cambiando la condición de 0 a 1. Apague el migrador Verts, que es el habilitado por defecto, a menos que desee utilizarlo.
- Guarde los cambios hechos al macro y luego cierre la base de datos.







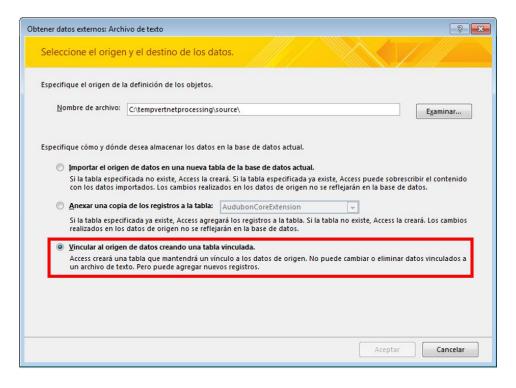
- **6.** Abra SourceTemplate-XXXX.mdb en Access (el que está en la carpeta "templates"), haga una tabla vinculada desde los datos fuente y llámela "Occurrence". Para ello, siga los siguientes pasos:
- 1. En el menú Datos Externos, importar datos (desde texto, desde Excel, desde Access, etc.). Si está importando desde Excel, se recomienda que primero exporte los datos desde Excel como csv. En la ventana que se abrirá, seleccione "Vincular al origen de datos".
- 2. En Examinar, busque el archivo fuente (los datos originales que copió a la carpeta "source" en el paso 2.). Haga click en Aceptar.
- 3. En la siguiente ventana, elija "Delimitados" para la delimitación, y "Primera fila contiene nombres de campos". Quizás deba también seleccionar un calificador de texto (" o ').
- 4. Si el archivo es .TXT vaya al menú Avanzado (en cualquiera de los paneles) y configure la Página de códigos a la codificación correcta. En general, la codificación es UTF-8, pero puede ser otra. Por favor compruebe la codificación de su conjunto de datos, mirando por ejemplo algunos registros que se muestran en la ventana de importación, y verifique luego de la importación examinando los datos. Casos en los que pueden aparecer errores son acentos y caracteres especiales, típicos de ciertos idiomas (e.g., ES, FR).



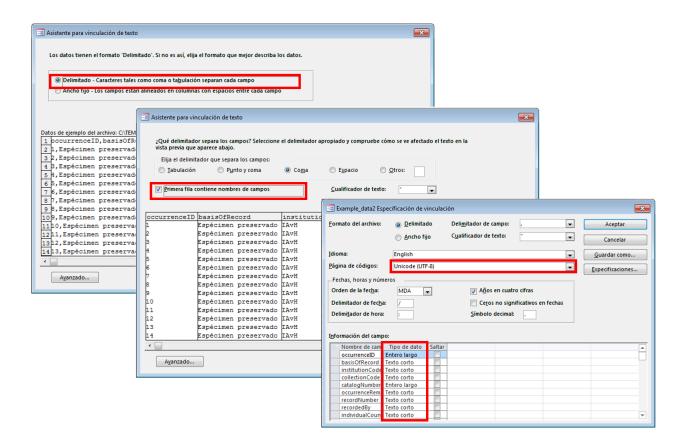


- 5. Configure para todos los campos de la tabla los Tipos de datos a "Texto corto", o "Texto largo" ("memo") si el campo puede contener muchos caracteres (i.e., más de 255; e.g., campos de comentarios, tales como georeferenceRemarks o locality deberían ser Texto largo).
  - 5. Recuerde nombrar la tabla vinculada como "Occurrence". Finalizar.

-NOTA: En algunos archivos, usualmente CSVs, el texto pueden haber "" alrededor del texto, y Access puede informar que hay problemas con algunos de los encabezados, lo cual resulta en el agregado de columnas innecesarias. Si sucede esto, debe cambiar el calificador de texto. Usualmente, puede ver si hay "" en los datos simplemente revisando el archivo en un editor de texto. Revise la tabla Occurrence para detectar diacríticos y otros caracteres especiales y ver si han sido reproducidos correctamente. Si no es el caso, o bien la codificación escogida es incorrecta, o bien los datos han sido sometidos a un historial de importaciones y exportaciones que no respetaron la codificación correcta.



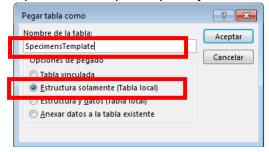




7. Elimine la tabla SpecimensTemplate. (Click derecho --> Eliminar)



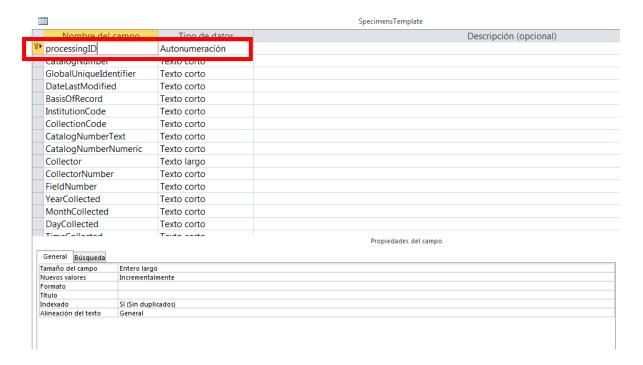
**8.** Copie la estructura (solamente) de la tabla Occurrence a una nueva tabla SpecimensTemplate. Para ello, seleccione la tabla Occurrence, haga click derecho y seleccione "Copiar", luego click derecho nuevamente y seleccione "Pegar", renombre como "SpecimensTemplate" y escoja "Estructura solamente".



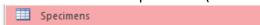
**9.** Edite la tabla SpecimensTemplate que acaba de crear: abra la tabla en vista diseño (click derecho --> vista diseño). Agregue una fila arriba de todo, nómbrela "processingID" y luego dele como tipo de datos 'Autonumeración'. Ahora haga de esta nueva fila una Clave principal: haga click derecho sobre el cuadrado gris junto al nombre del campo, y seleccione Clave principal. La fila tendrá ahora un ícono de clave principal (una pequeña llave). Guarde los cambios y cierre.







10. Elimine la tabla Specimens (click derecho --> Eliminar).



- **11.** Copie la tabla SpecimensTemplate a una tabla con nombre Specimens (sólo estructura, como hizo en el paso **8** con la tabla Occurrence, aunque dado que la tabla SpecimensTemplate está vacía, también podría escoger estructura y datos y sería lo mismo en este punto).
- **12.** Cierre la base de datos SourceTemplate-XXXX.mdb. Copie esta base de datos (toda la base de datos) desde la carpeta "templates" y péguela en la carpeta "workspace". Renómbrela como "Source-XXXX.mdb".
- **13.** En la carpeta "templates", renombre la base de datos modelo DwC2ExtractTemplate-XXXX.mdb para que coincida con el alcance del conjunto de datos para cual se realizará el migrador, donde XXXX es uno de "Audio", "Aves", "Eggs",

"Ent", "Fish", "Fossils", "Fungi", "Herps", "Inverts", "Mammals", "Plants", "Verts", etc. Si usa cualquier otro nombre de plantilla, tendrá que modificar el script "1a - RunMigrators.bat" para agregar un nuevo "CALL" para invocarlo, crear un script para ejecutar ese migrador usando el patrón "Dwc2migration-XXXX.bat", agregar consultas para los reportes en el macro "Do DwC2 Migration" y luego modificar la base de datos Aggregator Template.mdb para poder procesar los resultados.

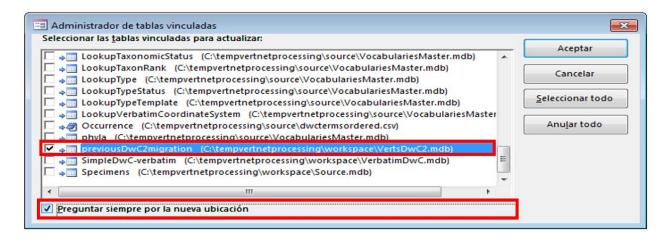
**14.** Abra DwC2ExtractTemplate-XXXX.mdb en Access (el de la carpeta "templates", que fue renombrado en el paso **13**). Allí puede organizar la barra de herramientas de la izquierda para ver los diferentes tipos de objetos (tablas, consultas, formularios, macros, etc.) y ordenarlos según diferentes criterios (tipo de objeto, fecha de creación, etc.). Puede seleccionar Todos los Objetos de Access para mostrarlos todos. Puede hacer click con el



botón derecho en una de las barras de sección, por ejemplo, Tablas, para ordenar por Nombre.

**15.** Enlace la tabla previousDwC2migración al migrador correcto (en la carpeta "workspace").

Para hacerlo, vaya a Datos Externos, haga click en Administrador de Tablas Vinculadas, desplácese hacia abajo hasta la tabla previousDwC2migration y seleccione la opción "Preguntar siempre por la nueva ubicación". Haga click en Aceptar. Se abrirá una carpeta para que elija un nuevo archivo para vincular a la tabla. Vincule su tabla a la base de datos apropiada en la carpeta "workspace" (por ejemplo, HerpsDwC2.mdb). A continuación, cierre la ventana Administrador de Tablas Vinculadas.



**16.** En esta base de datos plantilla (o modelo) en la que está trabajando, elimine la tabla vinculada denominada "Occurrence" (click derecho -> Eliminar), la que se ve así:



#### Occurrence

- **17.** Repita el paso **6** (el cual realizó en el SourceTemplate-XXXX.mdb) en esta base de datos (DwC2ExtractTemplate-XXXX). Es decir, cree una tabla vinculada a partir de los datos de origen y llámela "Occurrence".
- 18. Edite el contenido de la tabla LookupCollectionMetadata para que coincida con la colección que se está migrando. Para hacerlo, haga doble click en la tabla y realice los cambios respectivos en los elementos de la tabla de acuerdo con la información de la colección (metadatos). Confirme que el conjunto de datos no tiene una combinación de datos fósiles, preservados y de observación, ya que es posible que entonces necesite mapear el basisOfRecord de otra manera. Si todos los registros tienen el mismo basisOfRecord, puede capturarlo en esta tabla (LookupCollectionMetadata). Si no lo tienen, no lo mapee aquí y déjelo para el paso de mapeo más adelante.





#### Abra la tabla LookupCollectionMetadata

## LookupCollectionMetadata

- o Revise GBIF para institutionCode
- o Revise GRBio (<a href="http://grbio.org/">http://grbio.org/</a>) para institutionID use CoolURI si está presente, si no, use LSID institucional
- Revise GBIF para collectionCode
- o Agregue el link de la licencia Creative Commons apropiada en el campo license: CCO, CC BY, CC BY-NC.
- Agregue el enlace VN Norms al campo accessRights -<u>http://www.vertnet.org/resources/norms.html</u>
- o Agregue el collectionID de GRBio si está listado.

**19.** Si desea probar el migrador con una pequeña cantidad de registros (recomendado hasta que el migrador pueda ejecutarse completamente sin errores), use un filtro en Source para incluir sólo un número pequeño. Si el conjunto de datos es demasiado grande para ejecutarse a través del migrador (> ~ 1M registros), entonces puede ser necesario crear dos o más migradores, cada uno de los cuales utiliza un filtro en Source que acepta un subconjunto de los datos originales. Abra la consulta "Source" en vista de diseño y edítela para filtrar la tabla "Occurrence", si es necesario. Es decir: en la consulta "Source", verá que toda la tabla "Occurrence" está vinculada; si solo desea algunos de los campos incluidos, en vez de todos ellos, podría filtrarlos aquí. Los filtros pueden ser útiles por dos cosas:

FILTRO 1> en los campos. Por ejemplo, podría excluir los campos que sabe que los proveedores no estarían interesados en compartir (por ejemplo, los campos que capturan la ubicación de una muestra dentro de una colección - "número de cajón"). Esto no necesariamente tiene que suceder aquí, puede esperar hasta el paso de mapeo y no mapear esos campos.

FILTRO 2> en los registros. Por ejemplo, si tiene un conjunto de datos que es demasiado grande para procesar, puede cortarlo en partes y aplicar el proceso a los fragmentos por separado. Una forma de hacer esto es identificando un campo que pueda cortar fácilmente el conjunto de datos en fragmentos sin dejar ningún registro por fuera (por ejemplo, un campo de identificación numérico, que se rellena para cada registro: puede usar> # como primer filtro, operar, luego hacer lo mismo para el filtro complementario <= #).

NOTA: Si está trabajando con una base de datos grande, se recomienda que se establezca un filtro temporariamente, para que se ejecute el proceso en un subconjunto hasta que se esté seguro de que el migrador funciona. Esto le ahorrará mucho tiempo.





# C. Ahora es tiempo de mapear y personalizar el migrador, yuhuu!!

#### 20. Mapeo

Abra la consulta denominada "SimpleDwC-verbatim from Specimens" en vista de diseño (no la ejecute ahora). Esta será la plantilla para convertir Specimens en la primera versión de datos literales en SimpleDwC, es decir: el macro se ejecutará sobre esta consulta (paso 28).

En la vista de diseño de la consulta "SimpleDwC-verbatim from Specimens" verá algo como lo que se muestra a continuación:

Campo:	processingID	•	type	catalogNumber	Collector	Expr3: Nulo
Tabla:	Specimens		LookupCollectionMe	t Specimens	Specimens	
Orden:						
Anexar a:	processingID		type	catalogNumber	recordedBy	dynamicProperties
Criterios:						
0;		1)	2)	3)	4)	5)

Primera línea: campo en la tabla Ocurrence o en a tabla LookupCollectionMetadata, los que va a mapear.

Segunda línea: la tabla de la que procede cada uno de los campos de la primera fila (tabla Ocurrence o LookupCollectionMetadata). Esto estará en blanco si la primera línea contiene un valor constante o una expresión en lugar de sólo un nombre de campo.

Cuarta línea: los campos en los que se va a mapear.

Algunos ejemplos:

- 1) processingID. ¡Este campo no es un campo DwC! Sin embargo, es necesario para que el migrador conecte tablas durante algunos de los pasos. No lo toque. Deje el mapeo como está predeterminado.
- 2) Campos que se asignan directamente a un campo en la tabla LookupCollectionMetadata.
- 3) Campos que se asignan directamente a un campo en la tabla de Specimens.
- 4) Los campos que tienen un nombre no DwC, pero que se pueden mapear directamente (porque son equivalentes).
- 5) Campos que no se mapean directamente. Por ejemplo, en la tabla anterior no hay dynamicProperties en la tabla de Specimens, ni ningún otro campo con un nombre diferente que pueda mapearse exclusivamente a él. En estos casos, deje el campo nulo (Expr #: Null) o simplemente elimine ese campo (haga click con el botón derecho, "cortar") desde la vista de diseño. No elimine el campo de la tabla de Specimens.

El mapeo debe incluir sólo los campos que se pueden mapear directamente. Cualquier otro procesamiento debe ocurrir más adelante en el proceso. Por ejemplo, si no hay eventDate pero hay campos beginDate y endDate, no los mapee, hay una función específica para hacer eso más adelante en el proceso. Compruebe que la consulta esté visible sin errores intentando cambiar a vista de Hoja de datos (Datasheet view) sin ejecutarla (ya que eso intentaría anexar datos a la tabla Specimens, es decir: ino le haga doble click!). Si la consulta cambia a vista Hoja de datos (Datasheet view) sin errores, es lo que queremos en este momento. Tenga en cuenta que la vista Hoja de datos (Datasheet view) no mostrará ningún registro. Esto está bien.





Elimine todas las columnas que no apliquen a los datos (por ejemplo, contexto geológico para los no -fósiles). Esto es solo para mantener esta tabla limpia y simple (todos los términos DwC aparecerán de todos modos en el resultado final).

#### 21. "Ocultando" datos: restricciones\*

\* En este caso, se consideran restricciones ("encumbrances") al ocultamiento de ciertos tipos de datos. Por ejemplo, algún proveedor de datos puede no querer compartir la localización geográfica exacta de los registros.

En el macro "Do DwC2 Migration", edite y habilite las restricciones si es necesario. Esto está destinado a ocultar (retener) ciertos campos o registros que el proveedor podría no querer compartir. Hay restricciones para: registros, campos geográficos, campos de localidad y campos de georreferenciación. Se pueden crear también otras restricciones.

- Si no desea que ningún campo o registro esté oculto, configure la condición en el macro "Do DwC2 Migration" para "Withhold fields or records" a 0.
- Si quiere que algunos campos estén ocultos, entonces necesita editar la consulta de restricción correspondiente (cada consulta puede tener condiciones) y establecer la condición en el macro "Withhold fields or records" para esa restricción a 1. Por ejemplo: abra el consulta "Withhold geography fields" (click derecho -> Vista de diseño) y modifíquela según sea necesario.



#### 22. Anteponiendo toda la geografía literal al verbatimLocality

Este paso es para mantener toda la información geográfica literal "segura" en verbatimLocality. Recuerde que esto sucederá si más adelante (paso 28) en el macro "Do DwC2 Migration", habilita las consultas correspondientes llamadas "Prepend GGGG to verbatimLocality", donde GGGG corresponde con un campo de geografía superior, para hacer que verbatimLocality preserve toda la geografía superior original ya sea desde campos de Darwin Core o no (ver paso 28). Esto es necesario porque el migrador hará búsquedas en los vocabularios para estandarizar los campos de geografía de Darwin Core. Por ejemplo, "Prepend Waterbody to verbatimLocality" pondrá los datos del cuerpo de agua en el campo verbatimLocality, concatenados con cualquier otro campo que también esté anteponiendo. Como esto hará una concatenación, es importante tener los campos ordenados correctamente, del menos al más específico en el campo verbatimLocality, y por lo tanto en el orden inverso de las consultas, ya que éstas anteponen al verbatimLocality.

NOTA: los campos que está anteponiendo son los que están en los datos originales. Por lo tanto, tenga cuidado de usar los nombres de los campos que aparecen en esos datos



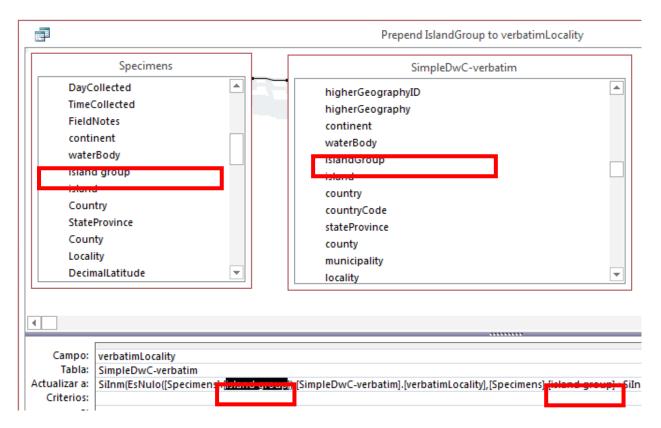


originales (tabla Specimens), y no los nombres que le daría después del mapeo. Ejemplo: supongamos que los datos originales tienen un campo "island group" (separado por un espacio en blanco), que se mapea al dwc:islandGroup. Para anteponer, use "island group", y no "islandGroup": abra la consulta correspondiente "Prepend islandGroup to verbatimLocality" utilizando la Vista de diseño y cambie el campo "Update To" para reflejar "island group" y no "islandGroup". Si el nombre del campo es el mismo que el nombre del campo Darwin Core, no tendrá que realizar ningún cambio en la consulta, ya que el campo que desea ya es el de la tabla de Specimens (por ejemplo, [Specimenes].[islandGroup]).

La expresión predeterminada se ve de la siguiente manera:

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

Recuerde cambiar el término cada vez que aparezca en la expresión.



Repita este proceso para cada campo geográfico, revisando los nombres de los campos en la tabla Specimens.

Agregue una nueva consulta para cualquier campo geográfico superior que no esté ya representado en la lista de anteponer (e.g., drenaje, parque nacional, etc.). Recuerde que si se agregan nuevas consultas, como una para Drenaje, debe agregarse al Macro (ver abajo, paso 28) antes de ejecutar el macro. Para agregar una consulta, copie / pegue / renombre una que ya tenga (en el panel izquierdo, haga click derecho en cualquier consulta previa ->





copiar, click derecho -> pegar). Edite la consulta nueva en la vista de diseño para que coincida con el campo correspondiente (por ejemplo, Drenaje).

#### 23. Poniendo toda la taxonomía literal en higherClassification

Edite la consulta llamada "Update higherClassification with verbatim taxonomy" para hacer que higherTaxonomy preserve toda la clasificación original. Para hacerlo, abra la consulta en vista de diseño. En el campo "Update To", la expresión predeterminada es similar a la siguiente:

```
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))
```

En esa expresión, tenga en cuenta que los nombres de los campos provienen de la tabla Specimens y, por lo tanto, deben tener los nombres reales que la tabla Specimenes utiliza para cada uno (por ejemplo, si el campo original es "Phyla" en lugar de "Phylum", debería llamarlo "Phyla", y esa parte de la expresión será: + Silnm (EsNulo (specimens.phyla), "", specimens.phyla)+

Saque de esa expresión todos los campos que no aplican para el migrador que está haciendo.

Agregue clasificaciones que no están representadas en la consulta, como suborden o subfamilia. Separe cada rango con " | " y vaya de menos específico (reino) a más específico (subgénero).

Mantenga todos los campos que correspondan a un rango de taxón hasta el subgénero (es decir, no incluya specificEpithet ni infraspecificEpithet, ni var, etc.). No incluya scientificName (o su equivalente).

Cierre la consulta y GUARDE los cambios.

#### Tratando con campos que no mapean directamente al DwC

**24.** Edite las consultas llamadas "LegacyXXX from Specimens" (donde XXXX es una de "Coordinates", "CoordinateUncertainty", "Dates", "Depths", "Elevations" o "ScientificNames") (estas son todas las consultas de anexado con un + verde en su ícono) para rellenar tablas Legacy Data (datos legados) con campos que no encajan directamente en SimpleDwC-verbatim o para campos que requieren un procesamiento especial (pasos múltiples o uniones a otras tablas) para transformar los datos originales en los campos y formatos recomendados en Darwin Core. Si se edita cualquier consulta Legacy, puede ser necesario actualizar el Macro para activarlas (vea el paso **28**).





- **25.** Agregue y personalice consultas para dynamicProperties si es necesario. Hay una serie de consultas "Update dynamicProperties for XXXX" ya disponibles, donde XXXX es la propiedad / característica (por ejemplo, medidas, plumaje, etc.). Si están presentes, actualice las consultas relevantes para reflejar los nombres de campo del conjunto de datos y para que coincidan con los campos de DwC apropiados en cada consulta. Si se utilizan cualquiera de las dynamicProperties, asegúrese de habilitarlas en el macro "Dynamic Properties" y de establecer la condición para ejecutar "Dynamic Properties" en 1 en el macro "Do DwC2 migration" (consulte el paso **28**).
- **26.** Edite la consulta "Update occurrence preparations for tissues" para que las preparaciones (preparations) incluyan información de tejido si no está incluído ya. Asegúrese de que los nombres de los campos coincidan con aquellos en los datos originales en el campo "Update To" de la consulta.
- **27.** Abra la consulta "Report missing catalogNumber" en vista de diseño y edítela para usar el campo de número de catálogo de los datos fuente.

#### Ahora es el momento de encender y apagar las funciones del migrador!!! :D

**28.** Abra el macro "Do DwC2 Migration" en modo de diseño (click derecho -> vista de diseño). ¡No lo ejecute todavía! (es decir, no haga doble click en él). Active y desactive los pasos según corresponda y edite las consultas que requieran personalización. Al encender / apagar, es posible que tenga que interactuar con tablas y consultas. La mayoría de los cambios ya los ha realizado en pasos anteriores, pero puede verificar a medida que enciende / apaga cada paso.

En el macro verá una serie de pasos, cada uno con una breve descripción de lo que hace, organizados en secciones. Cada paso se puede activar o desactivar cambiando la condición de 0 a 1 y viceversa. Puede ser útil mantener los datos originales abiertos en algún lugar para comprender lo que realmente contienen y cómo cada paso los afectaría. Si los datos originales no tienen algunas columnas particulares (por ej., Fecha), entonces todas las acciones llevadas a cabo sobre tales columnas deberían desactivarse (por ejemplo, todas las acciones sobre fechas), de lo contrario habrá un error porque la macro está tratando de hacer algo sobre datos que no existen.

Algunos pasos pueden tener una o las dos marcas siguientes en sus comentarios:

!!! significa que el paso es requerido y NO DEBE SER APAGADO.
\$\$\$ significa que la consulta puede requerir personalización y debe ser revisada su compatibilidad con la estructura y los contenidos de los datos fuente.

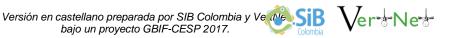
En algunos casos, ambos conjuntos de símbolos están presentes, lo que quiere decir que el paso es requerido y necesita personalización. Cualquier paso que no tenga \$\$\$ en su comentario no debe ser personalizado. Cualquier paso que solo tenga !!! no debe ser editado, ni apagado.





Algunos pasos que con seguridad se deben verificar / enmendar se enumeran a continuación:

- Sección de restricciones = macro "Withhold records and fields"
  - O El macro debe estar apagado a menos que haya restricciones. Las consultas en el macro deben estar activadas y personalizadas sólo si hay restricciones.
- Sección Anteponer = "Preprend..."
  - o Encienda todos los campos que están utilizados. Si, por ejemplo, Waterbody está presente en el conjunto de datos, la consulta "Prepend waterBody to verbatimLocality" debe estar encendida.
  - o Apague todos los campos que no son usados en el conjunto de datos, aquellos que no tienen una correspondencia en la tabla Specimens. Por ejemplo, si la tabla Specimens no tiene ningún campo correspondiente a la municipalidad (una subdivisión geopolítica de tercer nivel dentro del país, con ese nombre o cualquier otro), establezca la condición para la consulta "Prepend municipality to verbatimLocality" a 0. Si no lo hace, la macro parará con un error cuando se ejecute ese paso.
  - o Agregue una nueva acción en el macro para cualquier consulta agregada en el proceso. Ejemplo: si existe un campo de drenaje en los datos originales, se debería haber hecho una consulta "Prepend Drainage to verbatimLocality" y se debe agregar al macro. En este caso, dado que esto hará una concatenación, es importante hacer que los campos se ordenen correctamente, intente ir desde el menos específico (continente) hasta el más específico (en el macro, los verá enumerados al revés, el más específico en la parte superior). Para agregar una acción a el macro: copie y pegue una entrada de macro de la sección correspondiente. Edite la nueva entrada de macro para invocar esa nueva consulta.
- Sección Actualizar = "Update Simple DwC ..." Generalmente no se necesitan cambios para estas acciones, pero revise!
- Sección Legados (Legacy Sections). Habilite o dishabilite cada consulta emparejada
   "LegacyXXX from Specimens" y el macro asociado "Legacy XXX" de la siguiente lista:
  - o Coordinates Enciéndalo si el conjunto de datos incluye cualquier campo de coordenadas geográficas.
  - o CoordinateUncertainties Enciéndalo si el conjunto de datos incluye cualquier campo de incertidumbre de coordenadas.
  - Dates Enciéndalo si el conjunto de datos incluye cualquier campo de fecha del evento (year, month, day, eventDate, verbatimEventDate o equivalentes).
  - o Depths Enciéndalo si los campos de Profundidad están en uso.
  - o Elevations Enciéndalo si los campos de Elevación están en uso.
  - o ScientificNames Enciéndalo si un equivalente a scientificName está en uso y no hay campos separados para género y rangos menores.
- Sección DynamicProperties Encienda el macro "Dynamic Properties" si se capturará cualquier propiedad dinámica de los datos originales. Personalice y habilite cualquiera de dichos dynamicProperties en el macro "Dynamic Properties".





- Limpiar las secciones de la tabla SimpleDwC generic processing, por lo general sólo déjelo tal cual está, pero revise cualquiera que tenga \$\$\$ en el comentario.
- Sección Audubon Media Extension Encienda el macro "Audubon Media Extension" si hay enlaces de media (e.g., fotos, videos, sonidos) en los datos.
  - o Si hay datos de medios para ser procesados desde el campo associatedMedia en la Audubon Media Extension, active los pasos apropiados en el macro "Audubon Media Extension". Si se provee un archivo multimedia por separado, será necesario realizar una personalización adicional para vincular esos datos y rellenar la tabla "AudubonCoreExtension", incluyendo la modificación de la consulta "Append media to AudubonCoreExtension" en función del archivo vinculado.
- Sección RelationshipResource: un uso común de la extensión ResourceRelationship
  es notificar a GBIF que los identificadores únicos globales para los registros han
  cambiado. Por ejemplo, si hay un cambio en cómo se asignan los occurrenceID de
  una versión publicada anterior de un conjunto de datos,
  ResourceRelationshipExtension tendrá que rellenarse con los identificadores
  antiguos y nuevos. Active el macro "Resource Relationship Extension" si hay
  relaciones para agregar.
  - Si hay relaciones de recursos para agregar, active los pasos apropiados en el macro "Resource Relationship Extension" bajo "Add occurrenceIDs to RelationshipResourceExtension" en el macro.

# D. Ejecute el migrador! Por... primera vez!!!

29. Pruebe el migrador ejecutando el script "1a) RunMigrators.bat", en la carpeta "tempvertnetprocessing" con los CALL (llamadas) apropiadas habilitadas. Si la prueba falla, haga click en Detener / Cancelar. Importante: fíjese en el cuadro emergente qué consulta fue responsable del error y cuál fue la causa. La causa más común de errores es una referencia a un nombre de campo que no existe o es ambiguo porque existe en más de una tabla referenciada. La consulta que provocó el error tendrá que corregirse en la base de datos de plantilla (no en la base de datos temporal que está abierta mientras se ejecuta el .bat script). Cierre la base de datos y espere a que finalice el script.

### E. Sí... lo sé... errores... paró en la mitad del proceso, ¿no fue así?

**30.** Corrija el error que ocurrió en el paso anterior editando la base de datos de plantillas en la carpeta "templates" (no la base de datos creada en la carpeta "workspace" por el script "1a) RunMigrators").

Para hacerlo, abra templates\(specific Template) y busque la consulta que arrojó el error en el paso 27 (por ejemplo, si el error sugirió que specimens.eventDate no se encontró al ejecutar la consulta "LegacyDates from Specimens", abra esa consulta para verificar y actualizar el mapeo).





- Corrija el mapeo para los campos presentes el el conjunto de datos. Cierre la consulta y guarde.
- Una vez se han realizado las correcciones, intente hacer una prueba nuevamente empezando desde el Paso 29.
- **31.** Repita los pasos **29** y **30** hasta que se hayan corregido todos los errores.

#### F. Comprobando que todo salió bien

- **32.** Compruebe los resultados de la migración revisando los datos en la tabla SimpleDwC en la base de datos migrada en la carpeta "workspace". Confirme que todos los datos se hayan migrado a los campos correctos con las transformaciones correctas.
- **33.** Realice los ajustes necesarios para rellenar la tabla SimpleDwC de la forma más completa posible editando consultas y activándolas o desactivándolas en el macro "Do DwC2 migration" y cualquier otro macro necesario que éste invoque.
- **34.** Ejecute el script "1) RunMigrators.bat" con el CALL apropiado dentro de la activado (por ejemplo, "CALL DwC2migration.bat Aves") una vez más.
- **35.** Repita los pasos 32 a 33 hasta que todos los datos rellenen SimpleDwC de manera correcta.

#### G. Invocando múltiples migradores

**36.** A veces es necesario crear distintos migradores para los datos de distintas fuentes dentro de una institución y agregar esos conjuntos de datos en un solo conjunto de datos. Todos estos se pueden ejecutar juntos utilizando el archivo "1a - RunMigrators.bat" después de activar y desactivar los migradores apropiados como se describe en el Paso **4**.

#### **37.** Ejecute 1c - RunAggregators.bat.

NOTA: Realice este paso incluso si no realiza el Paso **36**. Se necesita para obtener reportes sobre issues (problemas) de línea vacías y para generar los archivos CSV que están listos para publicar en Darwin Core Archives.

- Cuando se haya completado, abra la carpeta ForIPT para encontrar los nuevos archivos txt que reportan sobre líneas vacías, codificación y otros problemas.
- Si, al revisar estos archivos txt, se reporta cualquier registro, éstos se deben reportar a la fuente para que puedan ser corregidos. Estos registros se habrán corregido en la migración, pero se deben corregir en la fuente.
- Luego revise la carpeta "Reports". Allí hay dos reportes que requieren de especial atención del proveedor de datos:





- o Informes de revisión-LinesWithNonprintingCharacters reporte cualquiera que aparezca a la fuente.
- o Reportes de Revisión-duplicate catalogNumber reportar cualquiera a la fuente si la fuente utiliza números de catálogo.

NOTA: la apertura de estos informes directamente con Excel no mostrará caracteres diacríticos u ocultos correctamente, por lo que deben revisarse con Notepad ++ o cualquier otro editor que respete la codificación y permita que se visualicen caracteres que no sean de impresión, si es necesario.

### H. Revisando y resolviendo vocabularios

- **38.** Revise los vocabularios. Para hacerlo, abra la base de datos "source\VocabulariesManager.mdb". Seleccione vista para ver todos los macros. Ejecute el macro "Unchecked Local Vocabs", que mostrará las ventanas para cada vocabulario que normalmente es rastreado y resuelto.
- **39.** Edite los vocabularios abiertos que muestran los valores que deben verificarse para mapearlos con las versiones estándar de los términos. Algunos de estos pueden saltearse, como preparations y lifeStage, que son increíblemente diversos. Estas vistas de vocabularios no se muestran por defecto en el macro "Unchecked Local Vocabs". Se pueden activar estableciendo la condición para ellos en 1 en el macro "Unchecked Local Vocabs".

# I. Ejecutando nuevamente el migrador para incluir los nuevos vocabularios resueltos

**40.** Ejecute de nuevo el script "1a - RunMigrators.bat" para hacer una copia de la tabla SimpleDwC con búsquedas de vocabulario incluidas.

#### J. Ensamblando los conjuntos de datos

- **41.** Ejecute el script "1b RunAggregators.bat" luego de que los migradores para todas las colecciones diferentes hayan finalizado hasta este paso para hacer un conjunto de datos migrados final agregados.
- **42.** Compare los recuentos de salida de SimpleDwCForVertNet.csv con el número esperado de las tablas SimpleDwC procesadas para la colección. Resuelva cualquier discrepancia.





## K. Pasos Finales: reportes y datos para publicación

- **43.** Archive los archivos de reporte de errores (llamados "Report \*" en la carpeta de reportes) para enviarlos al publicador de datos. Alerte al publicador sobre estos reportes para que puedan mejorar sus datos en la fuente, si lo desean.
- 44. Comprima y luego copie el archivo SimpleDwCForIPT.csv para enviarlo a revisión por el publicador de datos. Déjele saber al/los publicador/es que abrir el archivo directamente en Excel probablemente malinterpretará la codificación, que es UTF-8. Brinde orientación sobre cómo preservar la codificación al abrir en Excel (consulte http://vertnet.org/resources/downloadsinexcelguide.html para obtener ayuda). Cuando esté listo para comprimir los informes para su revisión, no necesita incluir "processed.csv" ni "LineWithNonprintingCharacters", pero sí incluya una copia apropiada del documento de Report Explanation (Explicación de Reporte).

45. Suba el archivo SimpleDwCForIPT.zip al recurso del IPT cuando esté autorizado a hacerlo por el proveedor de datos.

FELICITACIONES!!! :D :D

