
# POSA-offline: plant taxonomy and distribution

**ver 0.9 Apr 2008**



## 1. Introduction

The objective of POSA-offline is to make southern African plant information available in a format that can be incorporated into other databases. In particular, POSA-offline provides plant taxonomic and distribution data.

POSA-offline comprises a set of data extracts from the main PRECIS (PREtoria Computerised Information System) developed and managed by the South African National Biodiversity Institute (SANBI).

Among other directives, the Biodiversity Act (No. 10 of 2004) mandated SANBI to curate and disseminate plant taxonomic information for South Africa. POSA-offline is intended to support this mandate.

## 2. Terms of use

## 3. Updates

POSA-offline is updated on a monthly basis, and made available for download from the SANBI FTP site (ftp://ftp.sanbi.org) and the Plants of southern Africa website (http://posa.sanbi.org). Database keys are kept constant, so any particular item will retain its unique identifier from one monthly release to the next.

## 4. Package contents

The POSA-offline package comprises the following files:
POSAoffline_Families = List of plant families
POSAoffline_Genera = List of plant genera
POSAoffline_Species = List of plant species
POSAoffline_Synonyms = List of synonyms of plant species
POSAoffline_Species_GrowthForm = Species growth forms
POSAoffline_Species_Distrib = Species large-scale distributions
POSAoffline_Species_QDS = Finescale distribution records for plant species

Each file comprises comma-delimited entries with lines terminated with a carriage-return newline-character combination (\r\n). Field names are included on the first line of each file.

Scripts that may be useful to users of MySQL are contained in the mysqlscripts subdirectory, namely:
posaoffline_loaddata_example.txt = script file to load POSAoffline data
posaoffline_reloaddata_example.bat = sample batch file load of POSAoffline data
posaoffline_structure.sql = SQL script to build POSAoffline database

In addition, the package contains this manual, POSAofflineManual.pdf.

## 5. Fields and relationships

### POSAoffline_Families:

Fields:
FAMNO = Family number
Name = Family name
AltName = Alternative (or historical) family name
Notes = References and notes on the family
Compiler = SANBI compiler(s) of this family

Indexes:
FAMNO (PK)

Relationships:
POSAoffline_Families.FAMNO (PK) = POSAoffline_Genera.FAMNO (FK), enforced, one-to-many

Notes:
Compilers of families are subject to change. Contact SANBI to be put in touch with the current expert on a particular group.

### POSAoffline_Genera:

Fields:
GENNO = Genus number
FAMNO = Family number
Genus = Genus name
Author = Genus taxonomic authority
Compiler = SANBI compiler(s) of this genus

Indexes:
GENNO (PK)
FAMNO (FK)

Relationships:
POSAoffline_Genera.FAMNO (FK) = POSAoffline_Families.FAMNO (PK), enforced, many-to-one
POSAoffline_Genera.GENNO (PK) = POSAoffline_Species.GENNO (FK), enforced, one-to-many

Notes:
Compilers of genera are subject to change. Contact SANBI to be put in touch with the current expert on a particular group.

**POSAoffline_Species:**

Fields:
SPNO = Species number
GENNO = Genus number
SpeciesFull = Full species name
Genus = Genus name
Species = Species name
SpeciesAuthor = Species authority (taxonomic abbreviation)
Subspecies = Subspecies name (if applicable)
SubspeciesAuthor = Subspecies authority (taxonomic abbreviation)
Variety = Variety name (if applicable)
VarietyAuthor = Variety authority (taxonomic abbreviation)
OtherName = Other name (for forma, subvariety, etc.), this includes the taxonomic level
OtherNameAuthor = Other name authority (taxonomic abbreviation)
Min_Alt_m = Minimum habitat altitude (metres)
Max_Alt_m = Maximum habitat altitude (metres)
Min_Ht_m = Minimum plant height (metres)
Max_Ht_m = Maximum plant height (metres)
Lifecycle = Plant lifecycle
IsNaturalised = 'Yes' (naturalised) or 'No' (not naturalised)

Indexes:
GENNO & SPNO (PK) [note that this is a composite primary key composed of both fields]
GENNO (FK)

Relationships:
POSAoffline_Species.GENNO (FK) = POSAoffline_Genera.GENNO (PK), enforced, many-to-one
POSAoffline_Species.GENNO (part PK) = POSAoffline_Synonyms.CURGEN (part FK) & POSAoffline_Species.SPNO (part PK) = POSAoffline_Synonyms.CURSP (part FK), enforced, many-to-many
POSAoffline_Species.GENNO (part PK) = POSAoffline_Species_QDS.GENNO (part FK) & POSAoffline_Species.SPNO (part PK) = POSAoffline_Species_QDS.SPNO (part FK), enforced, one-to-many
POSAoffline_Species.GENNO (part PK) = POSAoffline_Species_Distrib.GENNO (part FK) & POSAoffline_Species.SPNO (part PK) = POSAoffline_Species_Distrib.SPNO (part FK), enforced, one-to-many
POSAoffline_Species.GENNO (part PK) = POSAoffline_Species_GrowthForm.GENNO (part FK) & POSAoffline_Species.SPNO (part PK) = POSAoffline_Species_GrowthForm.SPNO (part FK), enforced, one-to-many

Notes:
Blank fields may be interpreted as NULL (unknown) values, except for the various species name components (Species, Subspecies, Variety, OtherName), in which case they indicate that there is no value in the field.
SPNO values of 0 indicate that only the genus is know, e.g. “Protea sp.”

**PRECIS_online_Synonyms:**

Fields:
CURGEN = Current (non-synonymous) species genus number
CURSP = Current (non-synonymous) species species number
SYNGEN = Synonym species genus number
SYNSP = Synonym species species number
Reference = Literature reference which identifies the synonymous name with the current (non-synonymous) name.
SynSpeciesFull = Synonym species name in full
SynGenus = Synonymous genus name
SynSpecies = Synonymous species name
SynSpeciesAuthor = Synonymous species authority (taxonomic abbreviation)
SynSubspecies = Synonymous subspecies name (if applicable)
SynSubspeciesAuthor = Synonymous subspecies authority (taxonomic abbreviation)
SynVariety = Synonymous variety name (if applicable)
SynVarietyAuthor = Synonymous variety authority (taxonomic abbreviation)
SynOtherName = Synonymous other name (for forma, subvariety, etc.), this includes the taxonomic level
SynOtherNameAuthor = Synonymous other name authority (taxonomic abbreviation)

Indexes:
CURGEN & CURSP (FK) [note that this is a composite foreign key composed of both fields]
SYNGEN & SYNSP (PK) [note that this is a composite primary key composed of both fields]
SYNGEN (FK)

Relationships:
POSAoffline_Synonyms.CURGEN (FK) = POSAoffline_Genera.GENNO (PK), enforced, many-to-one
POSAoffline_Synonyms.SYNGEN (FK) = POSAoffline_Genera.GENNO (PK), not enforced, many-to-one
POSAoffline_Synonyms.CURGEN (part FK) = POSAoffline_Species.GENNO (part PK) & POSAoffline_Synonyms.CURSP (part FK) = POSAoffline_Species.SPNO (part PK), enforced, many-to-many

Notes:
The SYNGEN field is a non-enforced foreign key of the POSAoffline_Genera table because not all synonymous species genera will be reflected in the main Genera table. All the CURGEN entries will appear in the main Genera table.
Because a number of synonymous species names may be 'lumped' into a single currently valid species name, and conversely a single synonymous name may also be 'split' into a number of currently valid species names, the relationship between current species (PRECIESonline_Species) and synonyms (POSAoffline_Synonyms) is many-to-many. A single current species name may have a number of associated synonyms, and a single synonymous name may have a number of currently valid names.
SYNGEN and SYNSP will not appear in the POSAoffline_Species table (as GENNO and SPNO), so these keys are not directly useful but are included for completeness.

**POSAoffline_Species_QDS:**

Fields:
SPNO = Species number
GENNO = Genus number
GRIDREF = QDS (quarter-degree-square) (see notes)
MostRecentYear = Most recent year in which a specimen was collected in the defined QDS

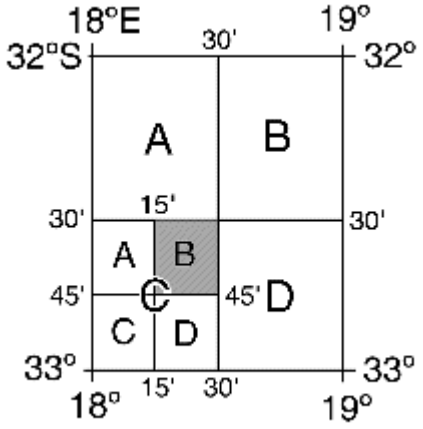
Indexes:
GENNO & SPNO (FK) [note that this is a composite foreign key composed of both fields]
GENNO (FK)

Relationships:
POSAoffline_Species_QDS.GENNO (FK) = POSAoffline_Genera.GENNO (PK), enforced, many-to-one
POSAoffline_Species_QDS.GENNO (part FK) = POSAoffline_Species.GENNO (part PK) & POSAoffline_Species_QDS.SPNO (part FK) = POSAoffline_Species.SPNO (part PK), enforced, many-to-one

Notes:
The GRIDREF contains QDS (quarter-degree-square) data in the form [xxyyab] where
xx = latitude (2 digits, assumed south of the equator)
yy = longitude (2 digits, assumed east of Greenwich)
a = A-D, degree quadrant, where northwest is ‘A’, northeast is ‘B’, southwest is ‘C’ and southeast is ‘D’.
b = A-D, quarter-degree quadrant (within degree quadrant defined by ‘a’), defined as for ‘a’.

A QDS corresponds to the area shown on a 1:50 000 map (15' x 15') and is approximately 27 km long (north-south) and 23 km wide (east-west).

See illustration below:




The shaded QDS is 3218CB.

**POSAoffline_Species_GrowthForm:**

Fields:
GENNO = Genus number
SPNO = Species number
GrowthForm = Growth form
Frequency = Frequency with which species is observed as possessing the growth form

Indexes:
GENNO & SPNO (FK) [note that this is a composite foreign key composed of both fields]
GENNO (FK)

Relationships:
POSAoffline_Species_GrowthForm.GENNO (FK) = POSAoffline_Genera.GENNO (PK), enforced, many-to-one
POSAoffline_Species_GrowthForm.GENNO (part FK) = POSAoffline_Species.GENNO (part PK) & POSAoffline_Species_GrowthForm.SPNO (part FK) = POSAoffline_Species.SPNO (part PK), enforced, many-to-one

Notes:
Definitions of growth forms may be found on http://posa.sanbi.org.

**POSAoffline_Species_Distrib:**

Fields:
GENNO = Genus number
SPNO = Species number
RegionCode = Region code
Region = Region name
CountryOrArea = Larger area name

Indexes:
GENNO & SPNO (FK) [note that this is a composite foreign key composed of both fields]
GENNO (FK)

Relationships:
POSAoffline_Species_Distr.GENNO (FK) = POSAoffline_Genera.GENNO (PK), enforced, many-to-one
POSAoffline_Species_Distr.GENNO (part FK) = POSAoffline_Species.GENNO (part PK) & POSAoffline_Species_Distr.SPNO (part FK) = POSAoffline_Species.SPNO (part PK), enforced, many-to-one

Notes:
South Africa is the most finely divided region. FSA, the Flora of southern Africa, comprises Namibia, Botswana, South Africa, Lesotho and Swaziland.

## 6. Examples

### Importing the data:

MySQL:

The mysqlscripts directory of your download contains the script posaoffline_structure.sql which will create the posaoffline database tables (note that the script was generated for MySQL v.5.0.45; prior versions (e.g. MySQL 4.0.18) may require minor tweaks to the script).

Due to some glitches in the text import process, you will notice that the fields Min_Alt_m, Max_Alt_m, Min_Ht_m, Max_Ht_m and IsNaturalised in table posaoffline_species and the IsNaturalised field in posaoffline_species_qds have all been set to VARCHAR instead of their correct numeric and Boolean field types. This is a known issue, but until it is resolved you may wish to change these fields to their correct field types manually once you have performed the data import.

In the same directory, the posaoffline_loaddata_example.txt file contains the commands needed to load the data. You will need to replace the text “C://directory//” in the file with the correct path to your download files. The file truncates the existing tables and reloads the contents of the delimited text files using commands of the form:

```
load data infile "C://directory//POSAoffline_Families.txt" into table
posaoffline_families fields terminated by ',' enclosed by '"' lines
terminated by '\r\n' IGNORE 1 LINES;
```

***Note the exclusion of the first line by using IGNORE, since this contains the field names.***

If you are working in an MS Windows environment, you can run the posaoffline_loaddata_example.txt file in MySQL with the batch file (posaoffline_reloaddata_example.bat) by double-clicking on it in Windows Explorer. Alternatively you can perform this action from the command prompt:

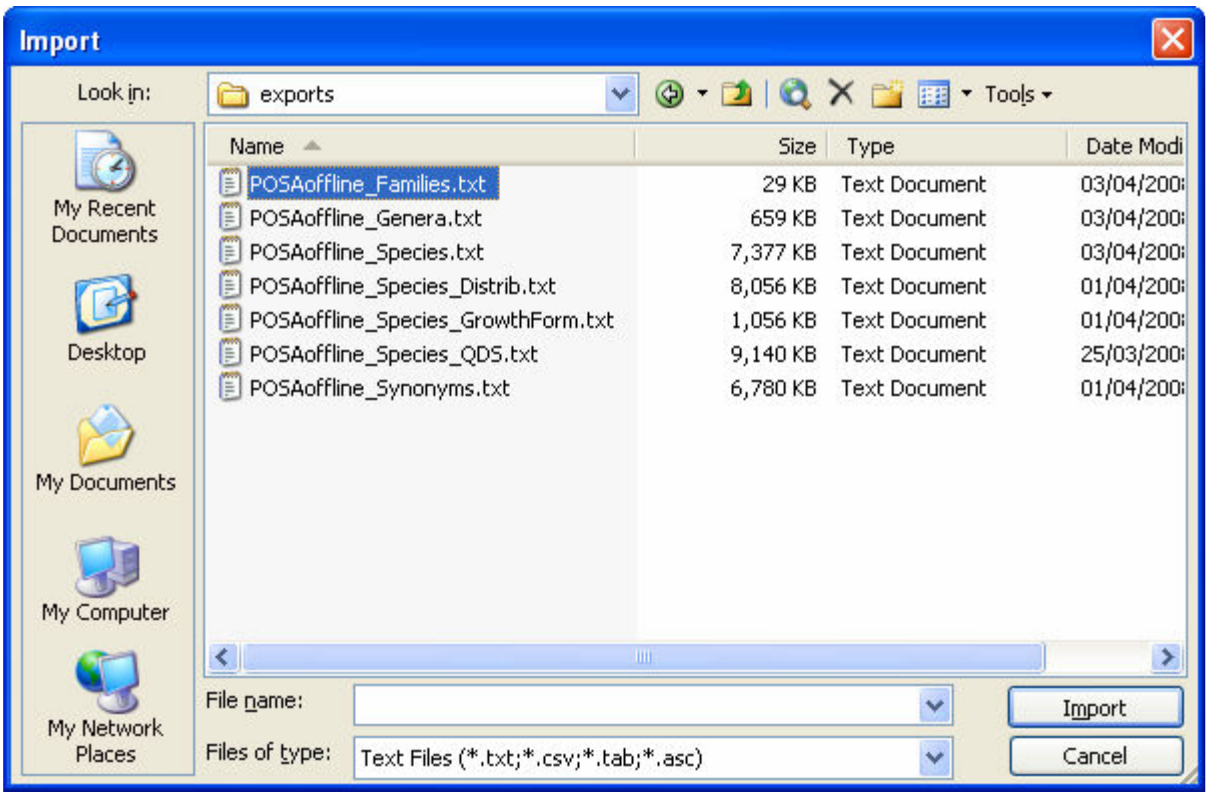
```
"C:\Program Files\MySQL\MySQL Server 5.0\bin\mysql.exe" posaoffline <
"C://directory//posaoffline//mysqlscripts//posaoffline_loaddata.txt"
```

Note that your MySQL executable may be in a different location and you will need to modify the path above accordingly.

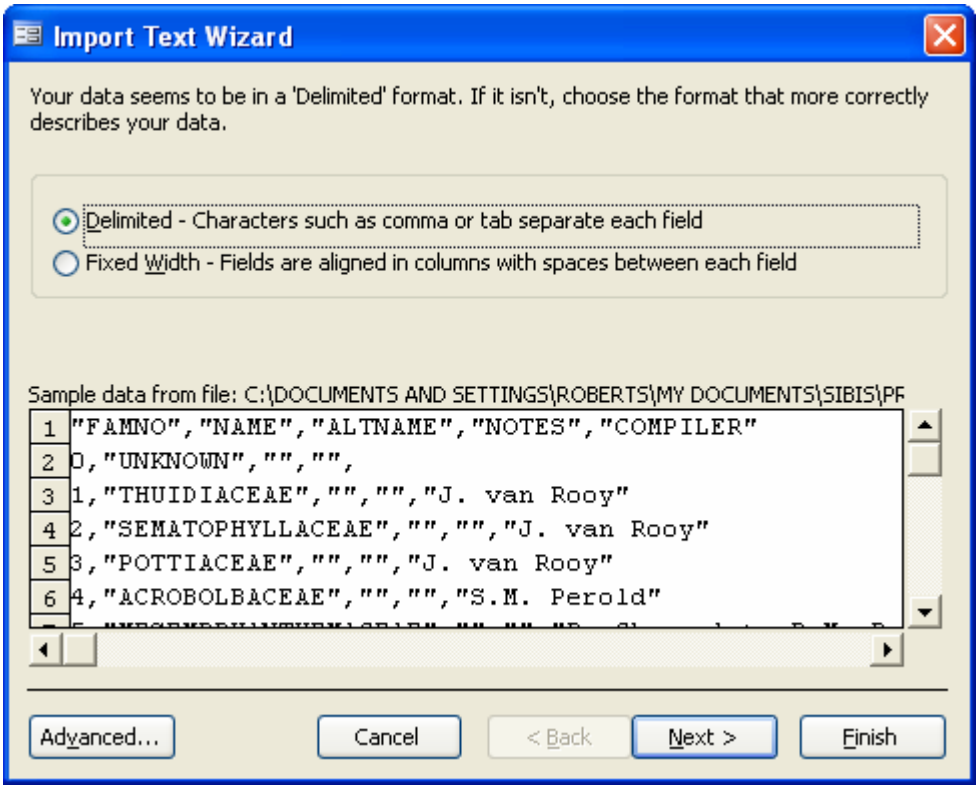
MS Access:

To import the data into an MS Access database, you may find the following steps useful:

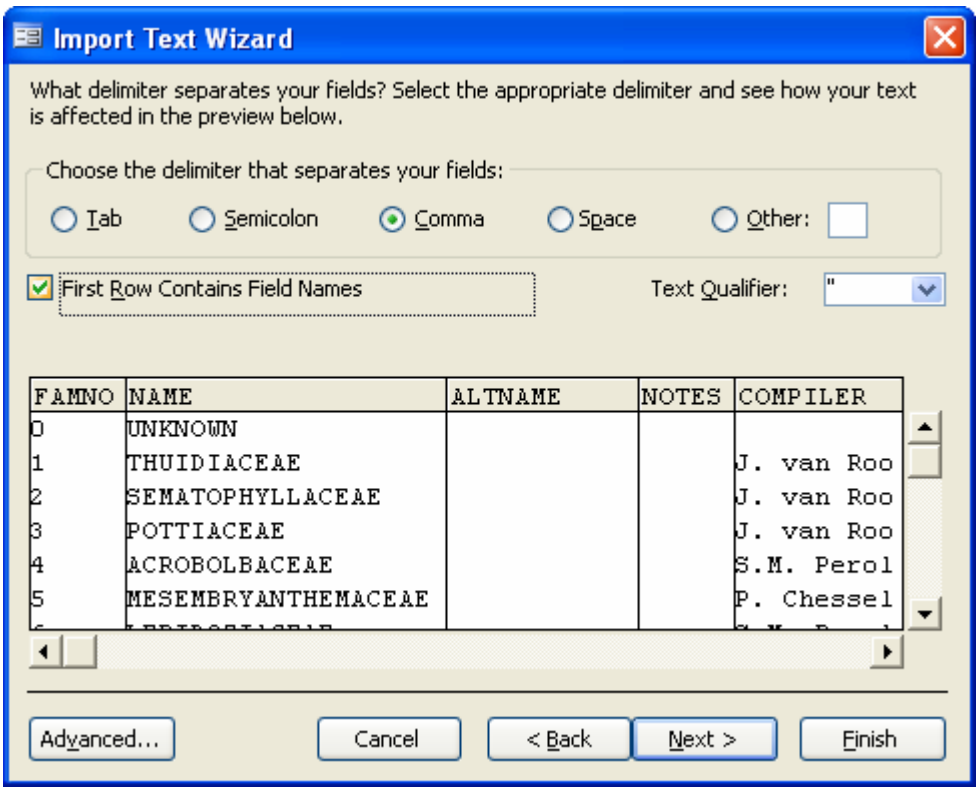
Create a new blank database. Then use *File -> Get External Data -> Import…* and navigate to the extracted POSA-offline files. Change the ‘Files of type’ setting on the window to view Text files and select the file you would like to import.



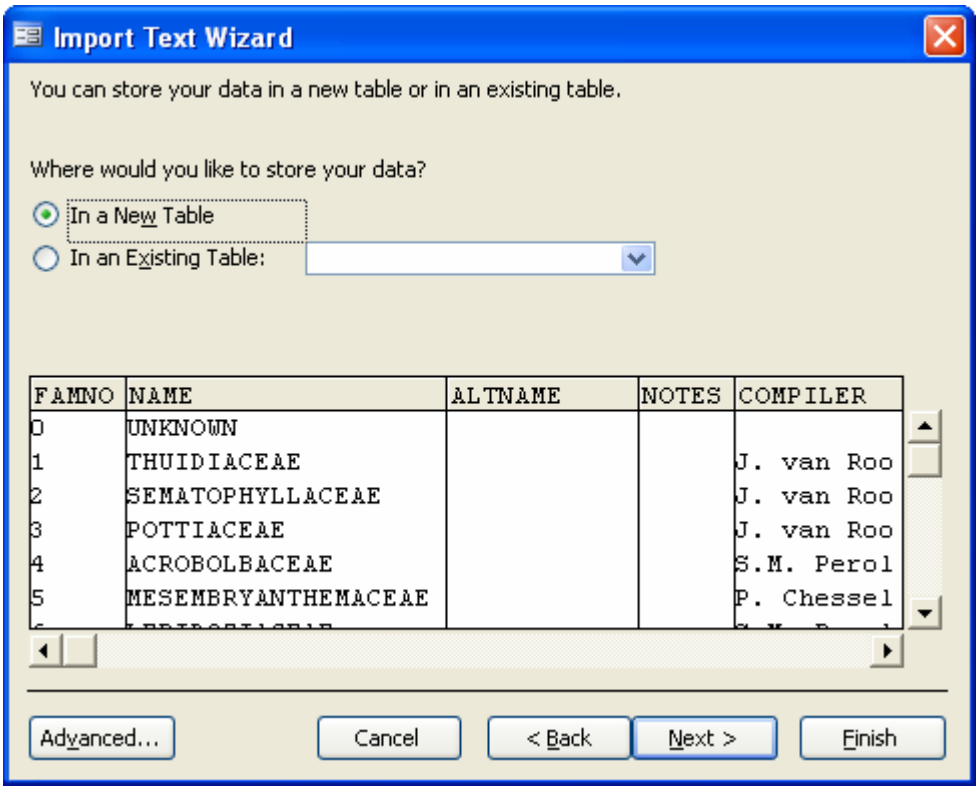

Leave the data as ‘Delimited’.




Check the box for ‘First row contains field names’.



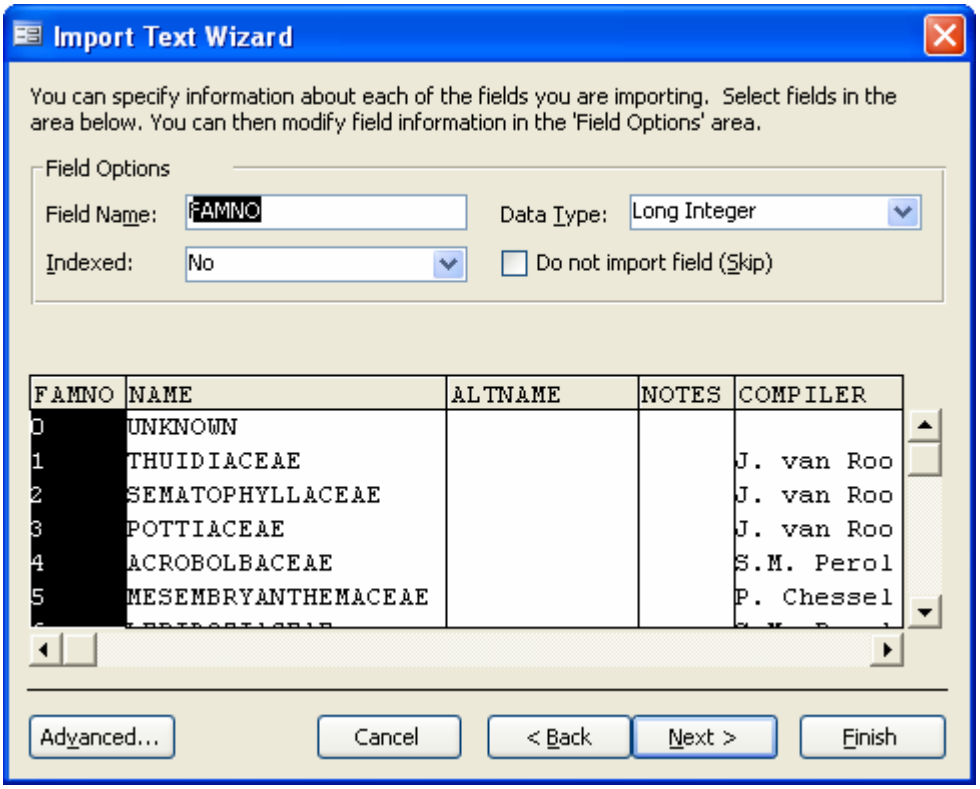

Import the data into a new table.



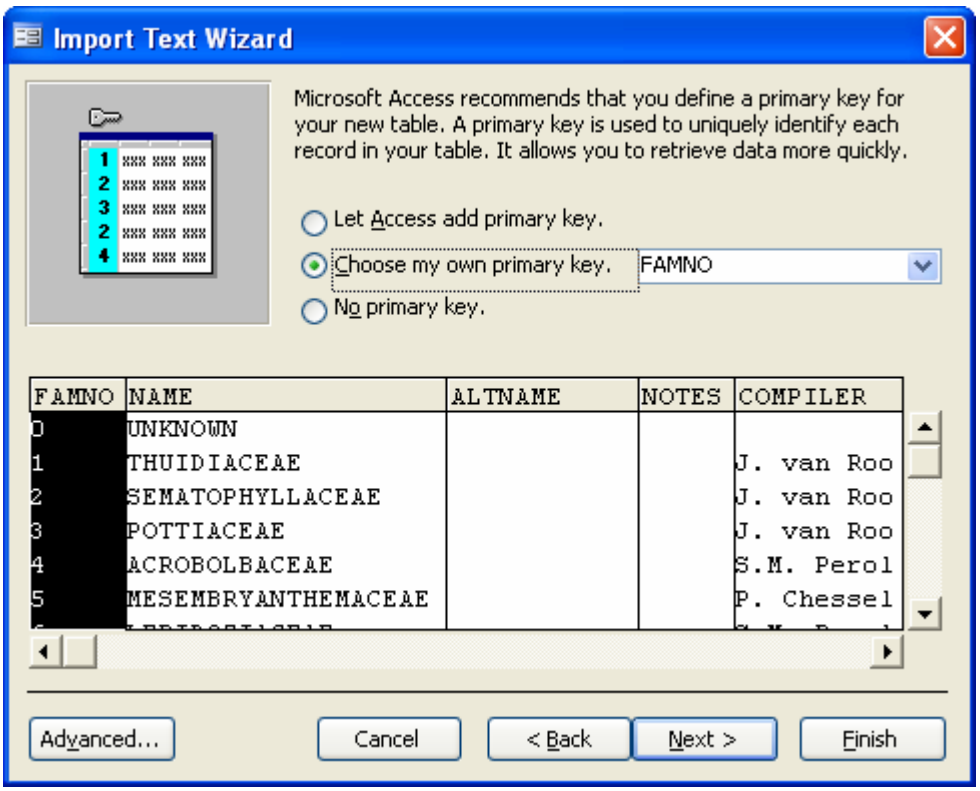

The field types are generally correctly identified.

For the POSAoffline_Species table, set the field types for Min_Alt_m and Max_Alt_m to Integer, those for Min_Ht_m and Max_Ht_m to Single and set IsNaturalised to Yes/No.

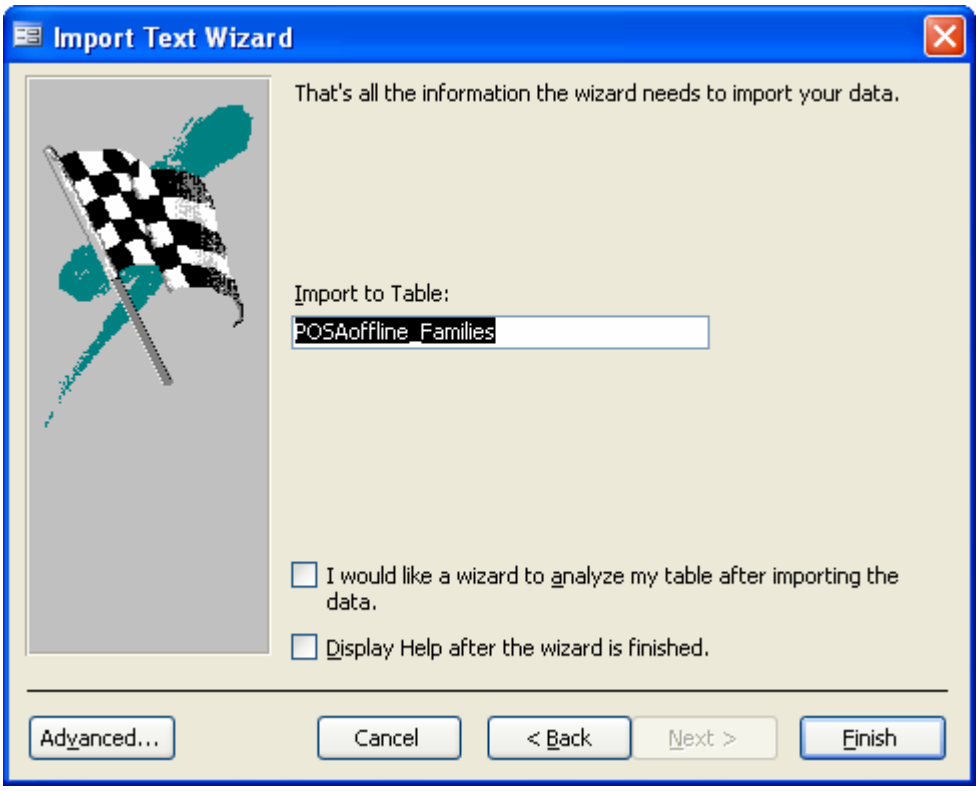
For the POSAoffline_Species_QDS table, set the field type for MostRecentYear to Integer.




Choose your own primary key, and set it to FAMNO (when importing the Species-related tables that do not have a single primary key field rather leave them with ‘No primary key’ and set it manually after the import process by selecting both GENNO and SPNO fields in the Table design view and clicking on the ‘Primary key’ button).




Save the table.



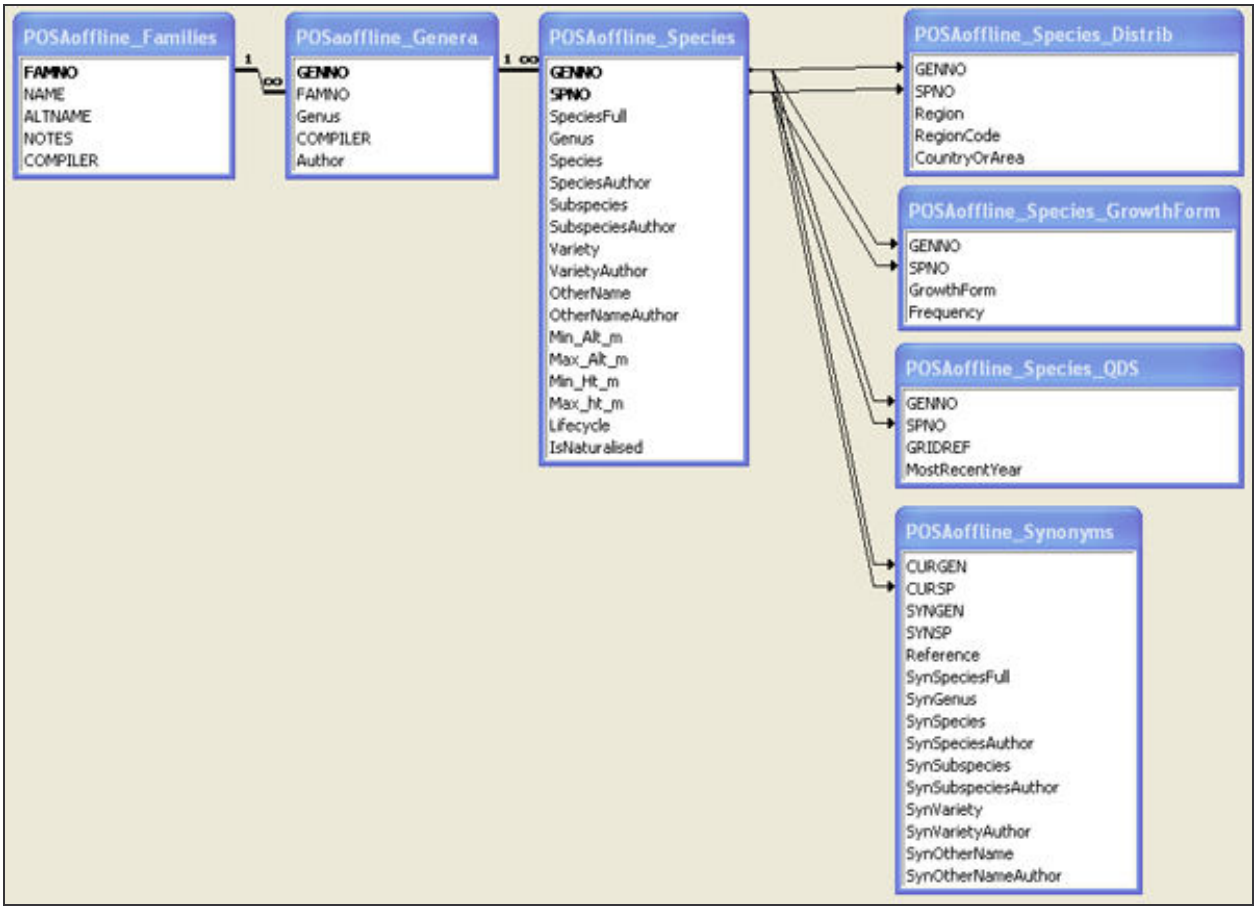

That's it! Repeat for all the text files. Note the field type manual settings described above, viz.:

**For the POSAoffline_Species table, set the field types for Min_Alt_m and Max_Alt_m to Integer, those for Min_Ht_m and Max_Ht_m to Single and set IsNaturalised to Yes/No.**

**For the POSAoffline_Species_QDS table, set the field type for MostRecentYear to Integer.**

## Querying the data:

The diagram below displays the relationships between the tables. To improve performance, it is worth indexing the GENNO and SPNO fields in each table, as well as the CURGEN, and CURSP fields in the POSAoffline_Synonyms table. Each of these will require a non-unique index.




Retrieving a list of true species within a particular family:
The following SQL will retrieve a list of all true species within the Aizoaceae and order them by species details:

```
SELECT POSAoffline_Families.NAME, POSAoffline_Genera.Genus,
POSAoffline_Species.SpeciesFull
FROM (POSAoffline_Families INNER JOIN POSAoffline_Genera ON
POSAoffline_Families.FAMNO = POSAoffline_Genera.FAMNO) INNER JOIN
POSAoffline_Species ON POSAoffline_Genera.GENNO = POSAoffline_Species.GENNO
WHERE (((POSAoffline_Families.NAME)="aizoaceae") AND
((POSAoffline_Species.SPNO)>0))
```

```
ORDER BY POSAoffline_Species.Genus, POSAoffline_Species.Species,
POSAoffline_Species.Subspecies, POSAoffline_Species.Variety,
POSAoffline_Species.OtherName;
```

Note that the filter for SPNO>0 will prevent the inclusion of “X sp.” records. In addition, the ordering could have been accomplished using the SpeciesFull field instead of each of the Genus, Species, Subspecies etc. fields, but in some instances this would produce an incorrect ordering (e.g. for subspecies and varieties).

Retrieving a list of annuals within a degree grid:
The following SQL will retrieve a list of all annual species (including ‘X sp.’ entries) that have been found within a full degree square (ie. any of the 16 QDS’s in that degree square).

```
SELECT POSAoffline_Species.SpeciesFull
FROM POSAoffline_Species LEFT JOIN POSAoffline_Species_QDS ON
(POSAoffline_Species.SPNO = POSAoffline_Species_QDS.SPNO) AND
(POSAoffline_Species.GENNO = POSAoffline_Species_QDS.GENNO)
GROUP BY POSAoffline_Species.SpeciesFull, POSAoffline_Species.Lifecycle, [GRIDREF]
Like "2518*"
HAVING (((POSAoffline_Species.Lifecycle)="annual") AND (([GRIDREF] Like "2518*")=-
1));
```

Note that the MS Access wildcard character ‘*’ has been used; on other systems you may need to replace this with ‘%’.

Retrieving a list of Proteas and their synonyms:
The following SQL will retrieve a list of all Protea species and their synonyms. The listing will not be unique with regard to the species listed, since a species having more than one synonym will be listed more than once.

```
SELECT POSAoffline_Species.Genus, POSAoffline_Species.SpeciesFull,
POSAoffline_Synonyms.SynSpeciesFull, POSAoffline_Synonyms.Reference
FROM POSAoffline_Species LEFT JOIN POSAoffline_Synonyms ON
(POSAoffline_Species.SPNO = POSAoffline_Synonyms.CURSP) AND
(POSAoffline_Species.GENNO = POSAoffline_Synonyms.CURGEN)
WHERE (((POSAoffline_Species.Genus)="protea"))
ORDER BY POSAoffline_Species.SpeciesFull;
```

Note that the LEFT JOIN allows those species which do not have synonyms to be included in the list.