Recorder 6

Linking Data held in *Recorder 6* with GIS

David Mitchel, exeGesIS, July 2009

Contents

1 Introduction

The object of this tutorial is to look at how to display data held in Recorder in GIS and load selections made in GIS back into the Recorder reporting wizard. There is no single solution as to how to link to GIS. It depends on your requirements, your GIS software, hardware and network specifications and the size of your dataset in Recorder so this tutorial will look at some of the issues to try to help you decide on the best method for you.

There are numerous GIS software programmes but the market is dominated by ESRI's ArcGIS (www.esri.com) suite and MapInfo (www.mapinfo.com). Other interesting and cheaper systems include Manifold (www.manifold.net) and the free Open Source programmes, QuantumGIS (www.qgis.org) and GRASS (grass.itc.it). This tutorial concentrates on ArcGIS and MapInfo.

2 Why link into GIS from Recorder?

Recorder does have its own internal mapping interface which allows the user to select data by polygon searches, so why link into GIS? The main advantages are:

- Data held in Recorder can be shared with staff via GIS who do not have access to Recorder itself

- Recorder data can be integrated with other corporate GIS datasets and systems for querying and reporting, for example including species records in planning system constraint checkers.
- The production of better maps showing species data
- Using GIS to perform complicated spatial queries (e.g. records within a certain distance of a river network, records on NE facing slopes above 200m, records within a set of polygons provided by a client) and pass the results back into Recorder for reporting or for the possible bulk updating of attributes like Altitude back in Recorder.

3 Making Recorder data GIS friendly

Before embarking on the process, it is important to consider a number of issues as these can affect which approach suits your purpose.

3.1 Issue 1: What type of data do you want to display?

GIS holds data as points, lines or polygons. A biological record held in Recorder is stored as a point record with the sample given a grid reference. Recorder, behind the scenes, stores the grid reference, its type (GB National Grid, Irish National Grid, Latitude & Longitude, UTM etc) and also its equivalent latitude and longitude. This is very useful as in GIS terms, an X and Y pair of figures are required for GIS to be able to plot a point whether they be Easting & Northing or Longitude & Latitude.

If you only have a grid reference like SO154339, as this is a not a separate X and Y pair, it needs to be converted into an Easting and a Northing. As the Longitude and Latitude figures are available in Recorder, you do not have to worry about converting to Easting and Northing. The only upshot of this is that currently (up to v6.14) when Recorder converts a grid ref to latitude and longitude it uses the OSGB36 datum but the projection of the point data uses the WGS84 datum which results in a slight discrepancy between the grid square data and the latitude/longitude point data (up to 200m at its worst). This issue is to be addressed in subsequent versions of Recorder 6.

For more information on the British National Grid and converting grid references to Easting and Northing, see www.ordnancesurvey.co.uk /gi/nationalgrid/nationalgrid.pdf.

If you do need tools to convert to Eastings and Northings automatically, JNCC have uploaded tools for Excel and Access to the Recorder website. See http://forums.nbn.org.uk/viewtopic.php?id=675.

However, a grid reference can itself have a number of meanings. If a record is given a grid reference of SO13, this is saying it is held somewhere within a particular 10km square. A grid reference of SO1533 says it is somewhere within a 1km square and SO154339 says it is somewhere within a 100m square.

Obviously the better the resolution, the more confidence you have of re-finding a record but in Recorder, many records are given grid references relating to the centroid of the site and so their accuracy can be spurious. Arguably if you are entering a species list for a square kilometre that you have searched, it is more accurate to give a 1km grid reference for the records as you are saying the observation was found

somewhere within this 1km square rather than a 100m square in the middle of this block where they were not actually found.

A problem arises when you plot these grid references in GIS as points as these points can give a false impression of accuracy. If you use the JNCC or Ordnance Survey tools for converting grid references to Eastings and Northings (or the Recorder lat/longs) and plot these in GIS, the points are plotted in the bottom left corner of the square they represent. Therefore a dot representing a record with a 10km grid reference will be plotted in the bottom left of that 10km square. This actual point may be outside the site boundary that the recorder said they were in or even in the sea but that is how it is plotted. If you then query these points against very detailed GIS datasets like SSSI boundaries, soils or geology, you can miss records or be told that some species are occurring on soil or rock types that are incorrect. Therefore there is a real danger in interpreting point records from Recorder from GIS and this must always be borne in mind.

The alternative is to plot the records as polygons with the size of the polygons relating to the resolution of the record. This is what the NBN Gateway uses and is actually a safer way to interpret the record set. On the negative side, it might mean that a query for records within a site will often pick up lots of records with poor resolution that were probably not found within that site but it is better to be presented with a large list and manually strip records out than miss key observations. This tutorial will look at generating both types of GIS dataset.

3.2 Issue 2: System Performance in GIS

Datasets held in Recorder can be enormous totalling millions of records at the LRC or National Scheme and Society level. Datasets of this size can be slow to load and query and you have to be much more careful about aspects like indexes and the number and type of attributes. Displaying subsets of the data rather than the full dataset can help, but inevitably, you will need to query the full dataset at times, so consider:

- indexing any attribute field in the GIS layer that you may be filtering on
- avoiding attribute fields that are memo or large text fields
- creating smaller subsets of the data possibly based on geographical area
- only displaying species of a particular status.

It is always possible to keep the GIS layer small in terms of attributes and relate or join this layer back into the main data table that stores all the rest of the attributes. This will also be looked at later.

3.3 Issue 3: Problem Fields

In Recorder, there are a number of fields that need to be interpreted before they make sense in a report or GIS table. These include dates, memos and varchars. Some fields in the Recorder reports contain spaces in the field names which can cause problems in GIS.

To allow the use of vague dates in Recorder, a date is held as a combination of three fields: Vague_Date_Start, Vague_Date_End and Vague_Date_Type. This allows the user to enter a date as “2009” or “Oct 2008 - Nov 2008”. In addition, in SQL Server, a date is not stored as “31/03/2009” but as “39903” which represents the number of days past 30 December 1899. So if a date is to be used in a GIS table, these two aspects have to be “translated”, but there are functions held in Recorder that help to transform the raw data into a meaningful date that can be used.

In the report wizard, some fields like *Associated Species* are created as a Memo fields which can mean when opened in ArcGIS, the word “blob” will appear for all entries. Fields like this have to be transposed into a varchar field in any query where they are used.

For performance reasons you may also want to avoid large text or varchar fields which can be created as part of the Snapshot as they canl slow any drawing or querying in GIS.These may be comment fields or fields like the Report Wizard *Sample Recorders* which have an 8000 character limit.

In the Report Wizard, some fields have spaces in their field name and this needs to be corrected for various GIS processes to work.

When you add a grid reference in the database, there is a field called Spatial_Ref_Qualifier which states how that grid reference was generated. This can be very useful in filtering records for meaningful GIS queries as it means that you can filter out grid references that were not the location of the actual observation (e.g. a site centroid). One problem with this field however is that if you are using the Recorder Import Wizard, any record being imported is tagged as “Import” rather than any meaningful so you just don’t know if these records are “good” grid references or just centroids.

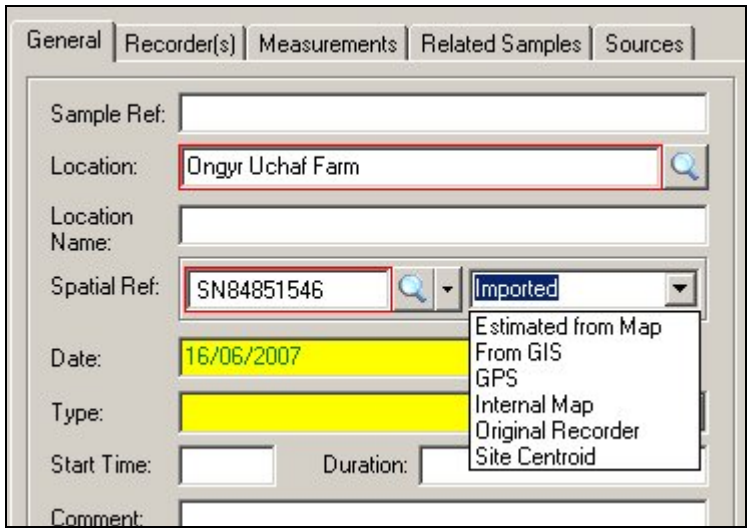


Fig 1: Spatial_Ref_Qualifier

Solutions to all of these issues are looked at below.

3.4 Issue 4: Flat file v relational tables

Recorder is a relational database with over 100 tables with many of the fields in any one table being a code relating to data held in another table. Hence any one table is usually not understandable on its own and displaying Recorder's Sample table in GIS would be fairly meaningless. GIS essentially stores its tables as flat tables (although it is possible to set up dropdown domains in geodatabases in ArcGIS) so the data in Recorder has to be transformed into a form that is readable in GIS. There are various ways to do this which we will now look at.

How to flatten the Recorder data structure

There are four main ways to prepare the data for GIS. They are:

- Using the Recorder Report Wizard and using the Export to ESRI Shapefile export option
- Using the Recorder Snapshot tool output
- Query, filter and transpose the Recorder Snapshot tool output and link to these results
- Creating a live link into Recorder using a SQL Server View

3.4.1 The Recorder "Export to ESRI Shapefile" option

This is by far the simplest way to get data out to GIS. Using the normal reporting wizard options, you can create a report choosing to filter the records or not, selecting which attribute fields to display and then when you have your result, just use the Export to ESRI Shapefile option in "Report Output".

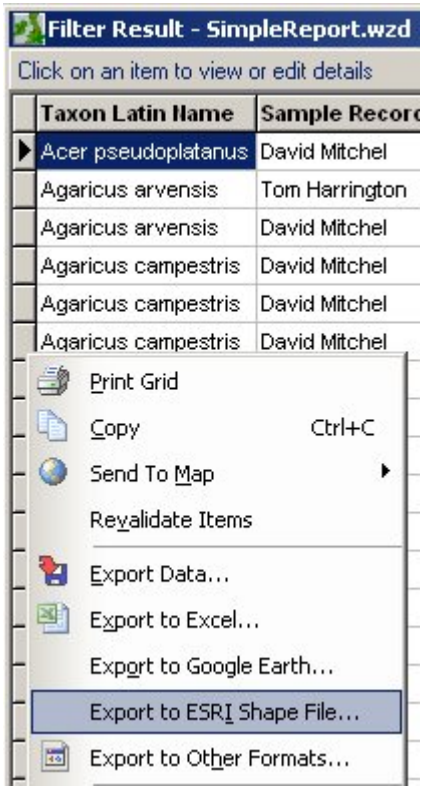


Fig 2: Export to ESRI Shape File

Advantages:

- Simplicity.
- Problematic attributes (see below) like *Date* and *Sample Recorders* are transposed into meaningful entries in the ArcGIS file automatically.
- a polygon dataset is created with polygons sized according to the resolution of the grid reference and this file can be opened directly in MapInfo in addition to ArcGIS.

Disadvantages:

- The process is more manual as this export option cannot currently be saved as a discrete file (combining both the report and export option) s you can with snapshots. This means the report has to be run each time and then the report output chosen and run, making the process more difficult to automate or schedule.
- You are confined to the fields offered by the Report Wizard and although these are now much improved, Sample.Spatial_Ref_Qualifier is missing which identifies whether the grid reference is a centroid, generated by GPS etc.

3.4.2 Use the Recorder Snapshot tool output

Using the Recorder snapshot tool is very simple. You set up a report as normal using the Report Wizard, choose your attributes and in the Report Outputs, choose "New Snapshot".

There are a few things to think about when choosing the attributes for this report. Although you choose the attributes from a list in the report wizard, the snapshot automatically adds more fields into the output. These are useful fields like:

- Taxon_Occurrrence_Key and Sample_Key so there is actually no need to choose "Obs Key" from the attribute list – it is already exported.
- If you choose "Sample Spatial Reference", Latitude and Longitude are automatically exported as well so there is no need to choose them again.
- Unfortunately, Sample.Spatial_Ref_Qualifier is not an option.
- "Taxon Group" and "Sample 10 km Square" are two very useful attributes when it comes to producing reports..

When the Snapshot Wizard is started, you need to logon with your SQL Server username and password and the Server name. These are the names set up when you installed Recorder. You will often have to type the Server name into the box as it may not appear on the dropdown list.

It is good practice to create a new database for your snapshot so it doesn't get overwritten by upgrades of Recorder. Manually type your new database name into the field and tick "Create Database if doesn't exist" and "Overwrite existing databases without prompting" so you can rerun the snapshot refreshing the data without numerous prompts.

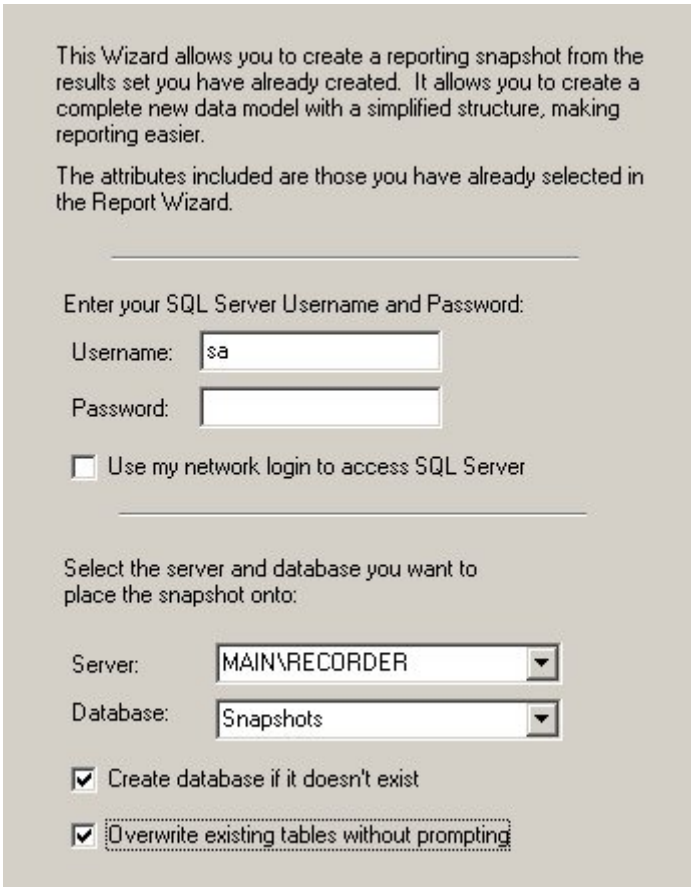


Fig 3: Taking a Snapshot 1

The next window (Fig 4) gives you the option to create a number of linked tables but it is simpler not ticking any of the options and just creating the one table. The following window (Fig 5) asks you to name the table created and again it is good practice to give it a new name each time as you may create a number of different snapshots.

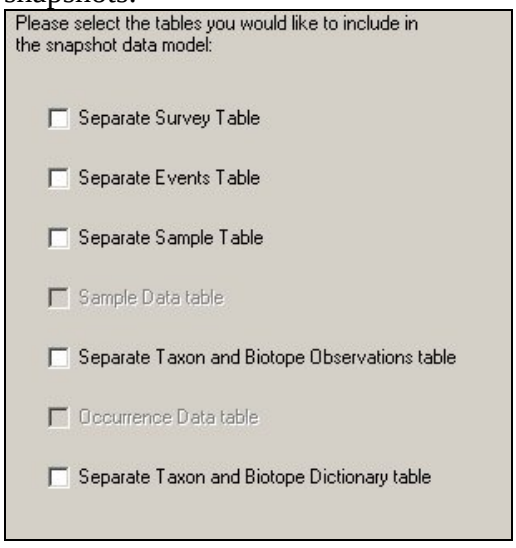


Fig 4: Taking a Snapshot 2

Fig 5: Taking a Snapshot 3

Click Run to create the database and when the database has been created, click "Close". You are then given the option to save the snapshot. Do this and also save the report you created using the wizard. This links the two files together and from now on, all you have to do to refresh the data in this snapshot is to go Reports – Run – Report Wizard Saved Reports and choose the report with the snapshot file name in brackets beside the report name.

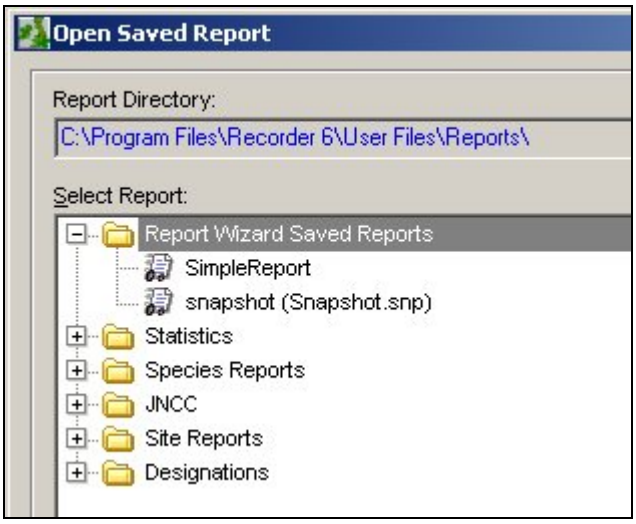


Fig 6: Re-running a Snapshot

3.4.2.1 The Snapshot Output

You could at this point, link this table into GIS and display the results. However there may be problems:

- Fields that have spaces in the titles which will cause problems in GIS queries.
- Dates presented as three fields (see above) and fields like Associated Species will all the entries shown as "blob".
- The table has no numeric indexed ID field which means that you can't create a live link into this table from MapInfo and also have to go through some extra steps in ArcGIS to make the records selectable.

To solve these issues, you need to create queries based on the snapshot table to make it more useful in GIS and examples of these are shown in the next section.

3.4.3 Query, filter and transpose the Recorder Snapshot tool output and link to these results

3.4.3.1 Creating Views in SQL Server – Software options

Recorder 6 stores its data in SQL Server. The default installation of Recorder is with MSDE which is the free version of SQL Server 2000. This has limitations on the number of users and database size so larger organisations may be using a full copy of SQL Server 2000 or SQL Server 2005. If you are running Recorder 6 on Vista, you will not be able to use MSDE and will have installed SQL Server 2005 Express which is also free.

MSDE and SQL Server Express 2005 do not come with a "front end" so to be able to look at and do anything on the back end of Recorder, you need a tool. You can download the free Microsoft SQL Server Management Studio Express

(http://www.microsoft.com/downloads/details.aspx?familyid=C243A5AE-4BD1-4E3D-94B8-5A0F62BF7796&displaylang=en) from the Microsoft website which allows you in at the database. To connect to the database, you need to know your server name and login again.

If you have full SQL Server 2000, to access the back end of Recorder, you will use Enterprise Manager. When creating views (or queries if you prefer this word) in Enterprise Manager or Management Studio Express, the interface is very similar. The examples below are all using Management Studio Express.

3.4.3.2 Creating a View

First, some terminology. In Microsoft Access and many other databases, you set up Queries on your data. In SQL Server, you set up ***Views***. They are essentially the same thing. On SQL Server you have just opened, you may have a number of databases listed. Apart from the system databases, you will have at least **NBNData** which is Recorder and your snapshot database. Tables, Views and Functions (in the Programmability section) for each database are listed in the tree below.

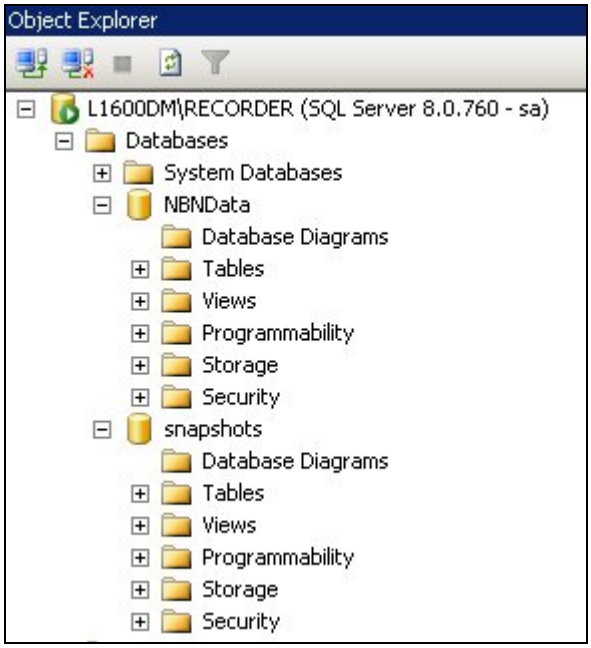


Fig 7: Management Studio Express Object Explorer layout

There are a number of stored functions in Recorder that can be used to clean up or transpose some of the fields (you can view them Programmability – Functions – Scalar-valued Functions – See Figure 8). These are particularly useful if you create your own raw views instead of using the snapshot tool and are all documented in the Network Installation Guide in the SQL Server Function section and include:

- ufn_RtfToPlaintext which strips off RTF tags in comment fields
- FormatIndividual which combines title, initials, first name and surname where appropriate
- ufn_GetFormattedReferenceName which formats a reference
- ufn_GetFormattedSpeciesName which combines the species name and authority

- LCReturnVagueDateShort llows you to convert the vague_date_start, vague_date_end and vague_date_type fields back into a meaningful date. E.g.

  NBNData.dbo.LCReturnVagueDateShort(Sample_Vague_Date_Start, Sample_Vague_Date_End, Sample_Vague_Date_Type) AS Sample_Date

Note that as these functions are in the NBNData database and not your Snapshot database. So to use them in a query in the Snapshot database, you have to state this when calling it by putting "NBNData" before the name of the function as shown in the example above.

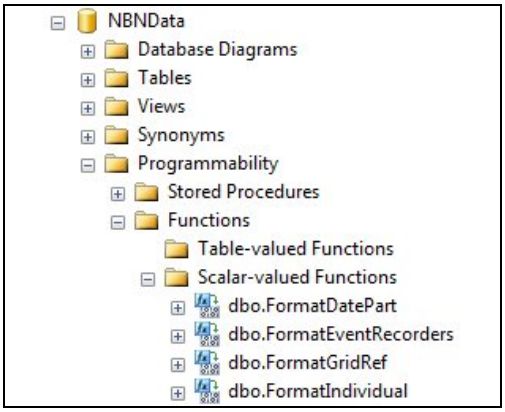


Figure 8 - Scalar functions that can be used to convert date (e.g. create a date from its 3 part fields)

It is also important to look at the properties of the fields in your snapshot table.

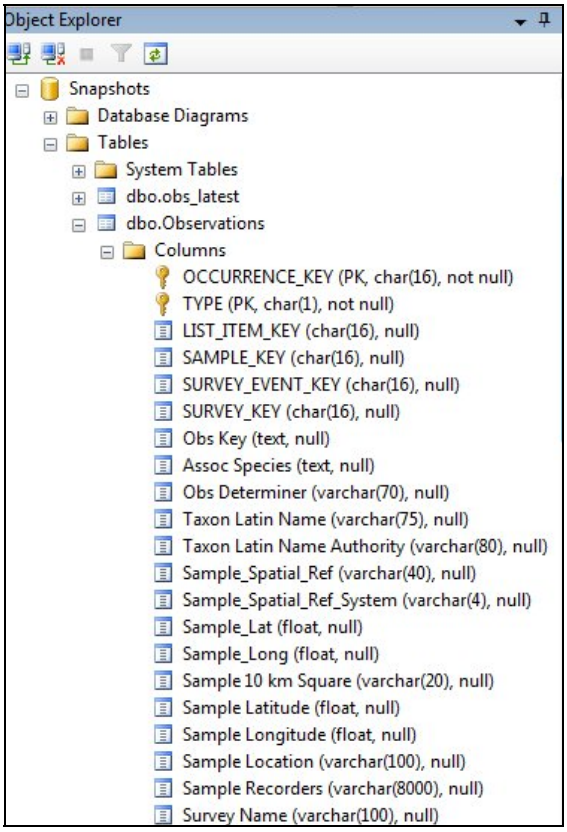


Fig 9: Tables, Columns and Properties Note the properties in brackets beside the column names.

- "Assoc Species" has properties of (text, null). This will lead to the word "blob" being put into the all the rows in the GIS table.
- "Sample Recorders" has a varchar limit of 8000 characters which is very long and will affect GIS performance.
- A number of the fields such as "Survey Name" have spaces in their field names which need to be fixed by using "AS" as shown in the example below.

To change the properties of the "Assoc Species" field, use the CAST function. The following SQL changes it to a varchar field with a 100 character limit and also renames the field removing the space in the field name. Note due to the space in the name, "Assoc Species" must be put within square brackets:

CAST([Assoc Species] AS varchar(100)) AS Assoc_Species

To create a new View, right click on "View" and choose "New View". Add the tables you want to load. For the Snapshot database, this will probably be only one table if you chose this option earlier when creating the snapshot. You can tick all the fields you want to display in the View and then edit the raw SQL in the window below. This is an example of the full SQL transforming the fields in my Snapshot (you might have different fields in your snapshot).

SELECT OCCURRENCE_KEY, CAST([Assoc Species] AS varchar(100)) AS Assoc_Species, TYPE, LIST_ITEM_KEY, SAMPLE_KEY,

SURVEY_EVENT_KEY, SURVEY_KEY, Sample_Spatial_Ref,
Sample_Spatial_Ref_System, Sample_Lat, Sample_Long, [Sample Location] AS
Location, [Obs Determiner] AS Determiner,
[Taxon Latin Name] AS Taxon_name, [Taxon Latin Name Authority] AS Authority,
[Sample 10 km Square] AS Sample_10_km_Square, CAST([Sample Recorders] AS
varchar(100)) AS Sample_Recorders, [Survey Name] AS Survey_Name
FROM dbo.Observations

As mentioned before, it can be useful to include the field Spatial_Ref_Qualifier so that you can filter to include only records that have been accurately georeferenced for detailed GIS queries (excluding centroid grid references). As this is not available in the snapshot output, you could create a new view on the raw Recorder data to pull this in. The following SQL would create a new View

SELECT SAMPLE_KEY, SPATIAL_REF_QUALIFIER
FROM NBNData.dbo.SAMPLE

This new view could then be linked into your previous View using the Sample_Key field to make Spatial_Ref_Qualifier available (you could also integrate this into the View above).

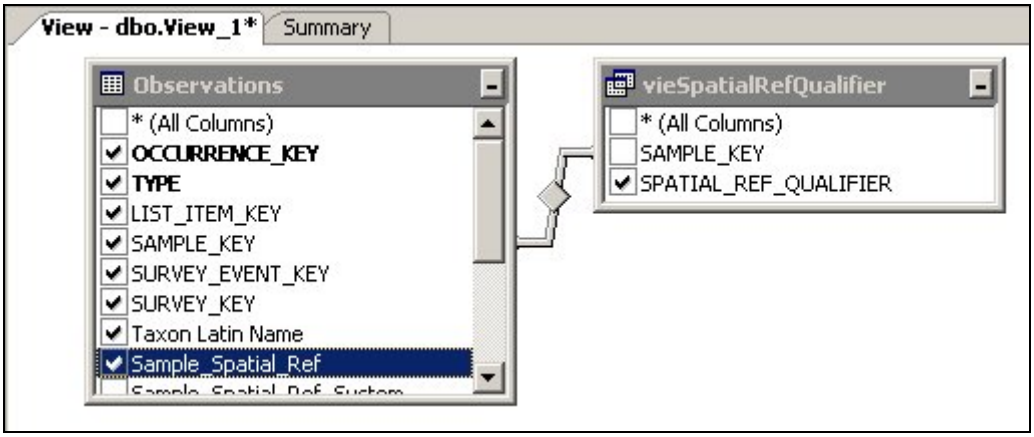


Fig 10: Joining tables or Views

As you get the hang of setting up views, all sorts of possibilities open up. One issue with displaying Recorder data in GIS is the potentially large number of records. By creating Views you can cut the dataset into manageable chunks and only display subsets of the dataset.

For example if your recordset is for a large area you could filter on Admin Area creating geographical subsets. Many users have a separate table of statuses for each species and if you use a table with all the statuses concatenated into one field, you could filter the records down to species with a specific status.

The good thing about this is to refresh the records, all you have to do is re-run the snapshot. Once the Views are setup, you do not have to change them as long as you are not adding new fields.

3.4.3.3 Indexes

Indexes are worth testing on your View if you are having performance problems. If you are regularly filtering on particular attributes, it is best to have this field indexed. These could be set up on the snapshot table say on species name. Running this after the snapshot was created could be scheduled as otherwise it is another manual task. An example of this is as follows:

```
USE [snapshots]
GO
CREATE NONCLUSTERED INDEX [name] ON [dbo].[Observations]
(
	[TaxonLatinName] ASC
) ON [PRIMARY]
GO
```

A very useful index would be to have a unique numeric index on the snapshot table as this will live selectable links into the Table or View from GIS. It is possible to add such a field onto the Snapshot table but this will be overwritten next time the snapshot is run.

Navigate to your Table in Object Explorer, open the tree up to the next level to show “Columns”. Right click this and add new column. Call it ObjectID, make it an Integer (int) and in the Column Properties, scroll down to Identity Specification and choose Yes for Is Identity.

The ideal would be for the Snapshot output to create this column, but this is not possible in the current version of Recorder.

3.4.3.4 Deleting tables or Views in the Snapshot database

As you might be creating Tables and Views in your Snapshot database regularly especially in the beginning, you will have to clear these out periodically. In Object Explorer, navigate to your database Table or View list. Right click the Table or View that you no longer want and choose Delete.

3.4.3.5 Updating

When the snapshot tool is run, a copy of the Recorder database is created. Any new records added to Recorder will not appear until the snapshot is run again.

3.4.4 Creating a live link into Recorder using a SQL Server View

You could create view(s) on the raw Recorder tables rather than use the snapshot tool. It would be better to do this from a different database than NBNData as you would do with the snapshots, so that any views are not overwritten by Recorder upgrades.

The advantage of this method is that this is a live connection and any record added in Recorder will appear immediately in GIS when the map is redrawn. However, there are risks in terms of database security in that records could be edited from GIS unless the View is read only, the setting up of which is beyond the scope of this tutorial.

Another disadvantage is that the views are more complicated to set up as you have to use the full NBN Data model rather than the flattened snapshot database. You have to weigh up the pros and cons of this method in your scenario before deciding if this is the route for you.

Performance in GIS may be an issue as the view would be quite complicated but this would depend on your local setup and would need to be tested.

4 Linking Recorder data in GIS

Now you have your views prepared, you can now view them in GIS. This tutorial is concentrating on ArcGIS and MapInfo only so we will now look at the methods used by both programmes.

With both systems, you must choose if you want to display the records as points or polygons representing the accuracy of the grid reference. If you are displaying points, you could create a live link into the View or create a new ArcGIS or MapInfo dataset from the View and work with that. The advantage of the latter is that performance will be better but the disadvantage is you are creating another step that has to be updated. If you are displaying polygons with sizes related to the accuracy of the data, this is by necessity an extra step.

4.1 ArcGIS

4.1.1 Live Link to the View

You first have to open ArcCatalog, go to “Database Connections” and double click “Add OLE DB Connection”. Choose “Microsoft OLE DB Provider for SQL Server” and Next. Add in your Server name, user name and password, select the database on the server and test the connection. The database name (e.g. “Snapshot”) will appear and by opening this, you will see the list of Tables and Views in that database. You can drag the database or table across into ArcMap and it will appear in the Table of Contents.

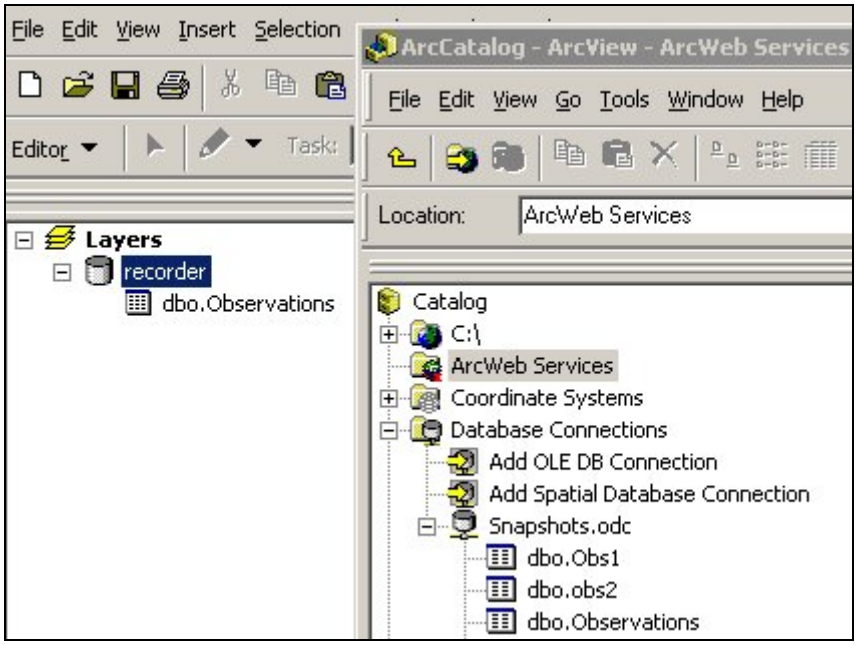


Fig 11: Loading a OLE DB Connection in ArcGIS

To display the points, you now have to tell ArcGIS what fields hold the spatial information. In this case it is Latitude and Longitude. Right click the table or View and choose "Display XY Data". Make sure the X field is Longitude and the Y field is Latitude and say that the projection is WGS84.prj (Geographic Coordinate Systems – World) as this denotes lat/longs.

The points will now be displayed but as you do not have an ObjectID field which is a numeric unique identifier, the points are not selectable.

If you now save the file as a shapefile, this field will be added and the points are now selectable.

If you would prefer to work with a live link into the Snapshot table saving a step in the process, you can use the Make Query Table tool in ArcToolBox (Data Management Tools – Layer and Table Views). The tool creates a query that includes a virtual ObjectID field thus making the points selectable. This layer is updated every time you open ArcMap. When using this tool, make sure that the ADD_VIRTUAL_KEY_FIELD option is chosen in the Key Field Options box.

4.1.2 Creating a new layer

When you have displayed the XY Data as described above, you can right click the layer, go Data – Export and export the data to a shapefile, personal geodatabase or file geodatabase.

After testing on a dataset with 1.4 million records, the time taken to redraw the data when held as a live link (including a Make Query Table layer), shapefile, personal geodatabase and file geodatabase was recorded.

Most of the formats performed the same (sluggish) taking up to 40 seconds to redraw but the best performance (10 seconds) was as a shapefile which is not surprising given

the simplicity of the file format. Performance therefore is the biggest reason for going this route and you have to make this decision based on your own local experience.

4.1.3 Performance tips

If you are still experiencing problems with performance, you could set up a view with a restricted number of attributes (e.g. the ones you need to filter or create thematic maps on).

Cutting out large text columns is a good idea. However these can still be made viewable by setting up a relate from your mappable ArcGIS layer to the main Snapshot table.

This means that this extra information is only viewed when you ask to show the data in related tables reducing processing time. To do this, the View or Table in the Snapshot database also needs an ObjectID field hence you may have to do a Make Query Table on this table as well. Once this is done, right click the layer that is displaying the Recorder data and choose "Joins and Relates". Set up a new Relate choosing the appropriate key to link the two tables on, e.g. Taxon_Occurrence_Key. To use the relate, when you have selected the records you want to view more data for, open the attribute table of the Recorder layer, click Options – Related Tables and the name of your relate. This will open another table with the extra data.

4.2 MapInfo

Linking to the tables in a Snapshot has been covered by Alan Hale on the Recorder Wiki. See **http://forums.nbn.org.uk/wiki/index.php/Using_the_Reporting_Snapshot_Tool_to_link_Recorder_6_to_MapInfo**. This tutorial will not duplicate this very good advice.

4.2.1 Live Link to the View

As Alan describes, this is not possible as there is no unique numeric ID in the Snapshot views unless you add one.

4.2.2 Creating a new layer

See Alan Hale's article.

4.2.3 Creating a polygon layer

There are two free MapInfo MapBasic tool (mbx) tools available to create a polygon layer from a point dataset with the size of the polygons related to the resolution of the grid reference. These are Alan Hale's tool (upload to NBN file uploads and link to it) and one from Tony Price of SERC (see http://juncus.org.uk/code/code.html). These are for use with GB grid references only.

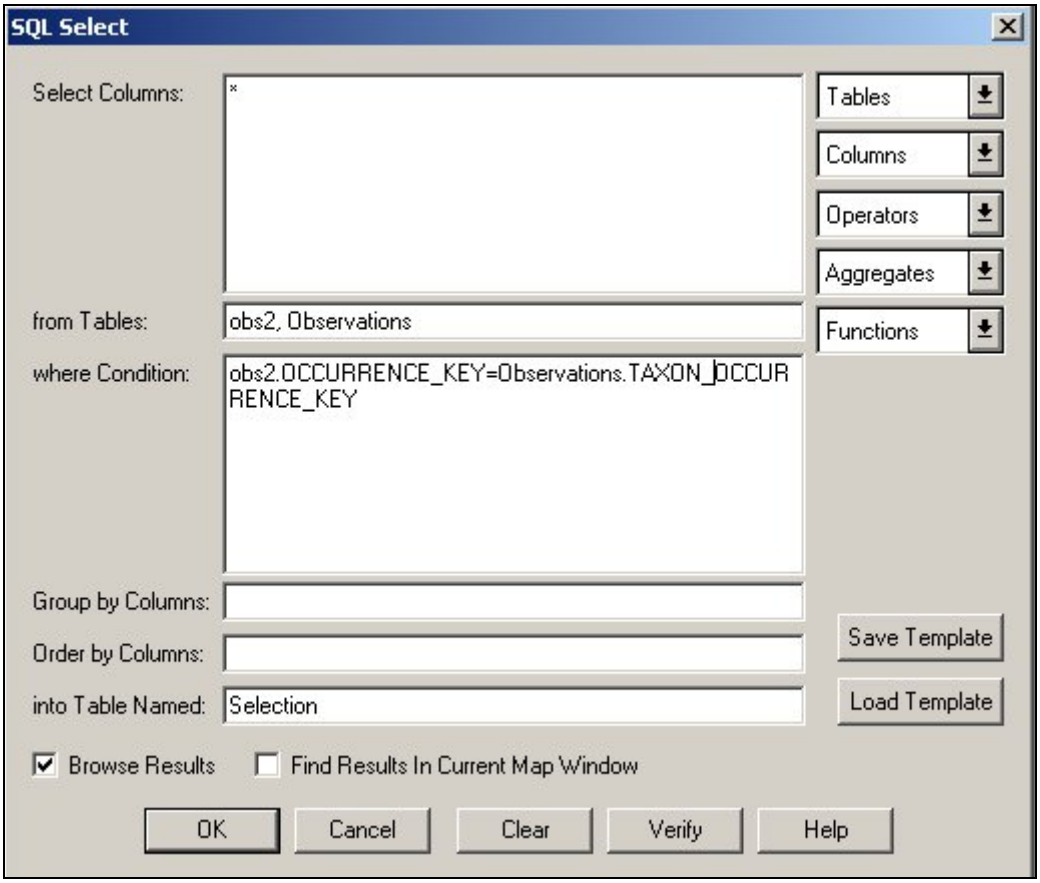
Alan's tool is an mbx and associated dll which must be saved to the same directory as the mbx. When you open the tool using Tool Manager, you are taken through a number of prompts locating the file to work on, identifying the grid reference field and setting the default styles for the different layers produced. Alan writes:

> “Takes as input a MapInfo, Excel or DBF table which includes a field for British OS grid references 'in either alphanumeric (e.g. SN123456), numeric (e.g. 22/123456) or tetrad (e.g. SN12J) format. Also 200m square grid-refs using a "tetrad" style lettering system ' e.g. SN4907K.
>
> The grid reference can contain spaces but not hyphens or other characters, except in the case of the numeric form, 'which MUST contain a forward slash.
> The grid reference (other than if in tetrad format) can be of any precision from 10km sq down to 1m sq precision '(i.e. 2-10 figure).
> The grid reference is passed to the GetCoords subroutine in the external library GetCord3.dll which returns Easting and Northing
>
> Note that the tool will cope with unpacked MapInfo tables, and very large tables (over 2,000,000,000 rows - in theory!)
>
> Version 12.1 adds a Bounds clause to the Create Map statement to give bounds which accord with CCW's standard settings 'of (0, 0) (2000000, 2000000).
>
> One thing that needs remedying is that it will not reject 5km squares expressed as e.g. SN44SE, but maps 'them as 1 km squares (out of range).”

When Tony Price’s tool is loaded, a new menu item is added “Make Region”. The file you are working on must not be open in the current workspace.

4.2.4 Performance tips

As with ArcGIS above, it may be useful to set up a join back into the Snapshot Table or View to access more detailed information about the record to help performance. To do this in MapInfo, in SQL Select, you set up a Join on the Occurrence_Key between the two tables. This creates a new query which you can display if necessary.



5 Things to do in GIS

When querying Recorder data against detailed GIS datasets, it must be remembered that the point for the Recorder observation may not really represent the location of the observation as a site centroid grid reference may have been used. This is why the ability to filter on Spatial_Ref_Qualifier is important and also if the data is to be used for constraint checking, it is safer to use the polygon representation of the data.

If you have a GIS layer with 10km squares as a polygon layer, there are some additional processes you can run on the dataset. You can create maps of recording effort, recording effort by species groups or even distribution maps all based on whatever mapping layers are available to you.

The first step in doing this is to have a 10km polygon layer. If you do not have one, in ArcGIS, this can be created using the free extension, Hawth's Analysis Tools (see http://www.spatialecology.com/htools/overview.php) and use the Sampling Tools – Create Vector Grid tool. In MapInfo, this is easily created using the MapInfo GridMaker tool. Once these layers are created, you have to attribute the polygons with the 10km name. Use the same format for the 10km name as used in Recorder Wizard attribute field for Sample 10 km square. Include this attribute and also the Taxon_Group attribute in the Snapshot output.

Now back in Management Studio Express, you can set up a new view summarising the records accordingly and then in GIS, you set up a Spatial Join on the 10km name field into the View. This will then allow you to create thematic maps with different colours representing recording effort, number of records, distribution maps etc. If you

created a point layer for the centroids of the 10km squares, the thematic mapping could be created on a point within the 10km grid. These are some SQL statements to create views on the main snapshot table:

Recording Effort

```
SELECT    COUNT(OCCURRENCE_KEY) AS NoRecords, [Sample 10 km Square]
FROM      dbo.Observations
GROUP BY [Sample 10 km Square]
```

Recording Effort by Taxon Group

```
SELECT    COUNT(OCCURRENCE_KEY) AS NoRecords, [Sample 10 km Square],
[Taxon Group]
FROM      dbo.Observations
GROUP BY [Sample 10 km Square], [Taxon Group]
```

Distribution Maps

```
SELECT    [Sample 10 km Square], [Taxon Name],
MAX(YEAR(Sample_Vague_Date_End)) AS LastRecord
FROM      dbo.Observations
GROUP BY [Sample 10 km Square], [Taxon Name]
```

The last view will give the year of the last record for all species within each 10km square. In GIS, you would then filter on a particular species with a thematic mapping set up to display its distribution.

6 Automating the whole process

One obvious issue about this whole process is the problem of keeping it updated. At the moment, there are a number of manual stages. Once the snapshot and views are set up, the manual stages are:

- Running the Snapshot
- Refreshing the link into the Views from GIS. With ArcGIS as you can set up a live link into the View using the Make Query Table, you wouldn't have to do any more but with MapInfo, the lack of numeric ID means that this is not possible and you will have to manually re-create the tab file when the Snapshot is run. Likewise with ArcGIS, if you are creating a shapefile from the link to the View for performance reasons, this will also have to be manually done.

All of this is possible to automate but it would not be using standard tools and needs development. Being able to schedule the Snapshot in Recorder would be a very useful addition to Recorder. In the meantime, it could be done from SQL Server but you could not use the Recorder snapshot tool and would have to create your own views to make a snapshot instead and then use SQL Server's scheduling tools to automate this.

This is not a problem if you are running full SQL Server but if you are running SQL Server Express, this may be possible using a batch file, Windows Task Manager or the SQL Server Service Broker. You first have to activate the Service Broker, e.g.

```
USE master;
SELECT name, is_broker_enabled FROM sys.databases;

USE master;
ALTER DATABASE snapshots SET ENABLE_BROKER;
```

Once this is done, the following website gives advice about setting up scheduling in SQL Server Express but it does need testing.

http://www.sqlteam.com/article/scheduling-jobs-in-sql-server-express

The second stage of automating the process refreshing the link into the View from ArcGIS or MapInfo can be setup with some macros or programming in ArcGIS and a MapBasic routine in MapInfo but these would need to be set up and are beyond the scope of this tutorial.

7 Passing GIS selections back into Recorder

If you have done various spatial selections in GIS, it can be useful to pass the results back into Recorder for reporting or for selecting records in NBN format. For example you may need to select all records within a set of National Trust polygons for selecting records to pass to them in Recorder format or you may be selecting records occurring on north east facing slopes for analysis purposes and need full access into the rest of the Recorder dataset.

This is possible by using CSV files (outputted from the resulting GIS query) in XML reports. For more information on this, see this article on the NBN forum: http://forums.nbn.org.uk/viewtopic.php?id=81

This same route could be possibly be used to bulk update measurement records like altitude, soils and geology which can easily be determined in GIS for the appropriately geo-referenced records. This would need more investigation but it could certainly be done in pure SQL. There are a number of issues about data security so this would need to thought about carefully but this route would be more preferable to doing it all by hand in Recorder!!