

## ArcGIS aquatic carbon database: guidance notes

The information contained in the master Excel spreadsheet has been incorporated into an ArcGIS relational geodatabase. This allows the data to be viewed and queried via a map interface, and also splits the large and cumbersome Excel table into a number of smaller, more manageable sub-tables which are linked together by *relationships* (explained below). This offers significant benefits compared to the Excel version, but it also makes accessing and manipulating the data more complicated – especially if you're unfamiliar with ArcGIS.

This document is designed to introduce the specifics of the aquatic carbon geodatabase. If you've never used ArcGIS before it is strongly recommended that you learn the basics before starting here; providing a complete beginners' course in ArcGIS is beyond the scope of these notes.

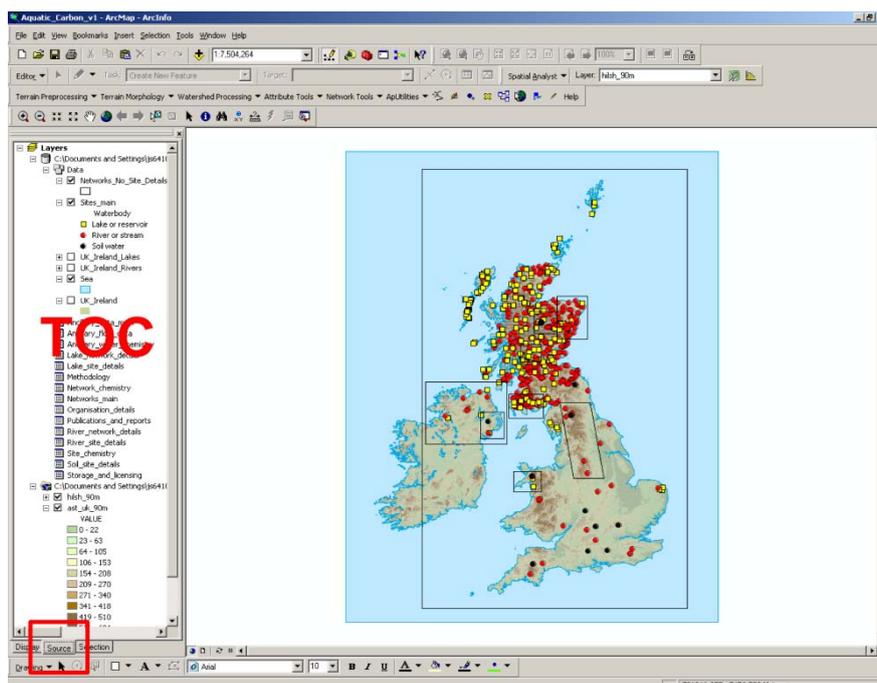
Please also note that, in order to provide maximum flexibility and to allow for future development, the user has complete and unrestricted access to all aspects of the geodatabase. This means that you can re-structure things, add or delete records and build very complex queries, but you can also remove things accidentally or create false relationships – please use with care!

As far as possible the database structure is generic, so anyone with experience of ArcGIS geodatabaseing should be able to expand or adapt the database using standard methods.

### Part 1: Copying and opening the geodatabase

Please note the following important points:

- The map document and all the data **must** be kept together with their current structure. If you need to move the database to a new location (e.g. to a hard-drive or memory stick), be sure to copy the **whole folder** "*Aquatic\_Carbon\_Database\_v2*". Do not alter or separate the contents of this folder unless you know how to re-connect the data links in ArcGIS.
- To open the database, double-click the ArcMap document called *Aquatic\_Carbon\_v2.mxd* (if you're using ArcGIS 9.3 or 9.3.1) or *Aquatic\_Carbon\_v2\_9-2.mxd* (for ArcGIS 9.2). Both located in the folder *Aquatic\_Carbon\_Database\_v2*.
- Once ArcMap has opened, click the **Source** tab towards the bottom left of the ArcMap screen (marked by a red box on the image below). **Always do this** – it's the only way to get access to the data tables stored in the database.



## **Part 2: The basics**

If you're not used to working with ArcGIS, the following tips may help:

- You will need to be familiar with the ArcMap *Tools* toolbar, especially the *zoom in*, *full extent*, *identify* and *select features* tools. This link shows what these tools look like and explains what they do: [http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=ArcMap\\_Tools\\_toolbar](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=ArcMap_Tools_toolbar)
- The column down the left hand side of the screen is called the *Table of contents* (marked TOC on the previous image). Having clicked on the *Source* tab as mentioned in Part 1, you should see a list of all the feature classes and data tables in the map. Clicking the check-boxes beside each layer's name allows you to turn them on and off.
- Layers in the map have tables (called *attribute tables*) associated with them. To open the attribute table for a layer, right-click the layer's name in the TOC and choose ***Open attribute table***.

Attribute tables store all of the information that ArcGIS needs to locate and symbolise different map elements correctly, as well as any additional information that users wish to add. Each row in the attribute table corresponds to a single *feature* (point, polygon etc.) on the map. Each column of the attribute table represents a different property that features in this attribute table can have.

Unlike graphics packages such as Illustrator or Inkscape, the appearance (colour, shape etc.) of features on the map is controlled by entries in the attribute table. For example, to determine whether a site should be drawn with a yellow square, red circle or black circle, ArcMap looks in the *Waterbody* column of the *Sites\_main* attribute table. Entries with "Lake or reservoir" are drawn with yellow squares etc.

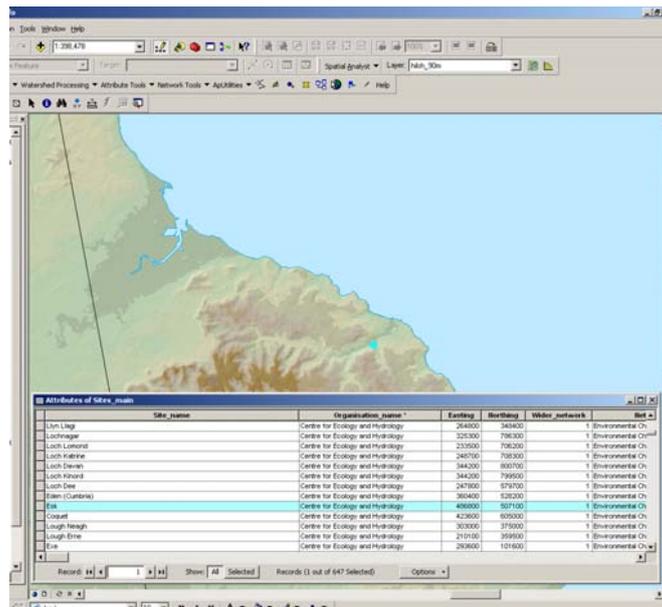
- In the TOC you should also see a number of standalone *data tables*, beginning with *Ancillary\_data\_main* and ending with *Storage\_and\_licensing*. To open these data tables, right click the table name as before, but this time choose ***Open***.

These data tables are just like attribute tables, with the crucial difference that they do not contain any spatial information. Attribute tables always have a column called *Shape*, which stores all of the information needed to locate each feature on the map. Data tables do not have a *Shape* column: they store additional information that is *related* to features on the map, but they do not directly influence where or how features appear.

- You will need to be able to *select* objects in order to find out about their properties. This can be done in two ways:
  - i. **By clicking on the map.** First choose the *Select features* tool from the tools toolbar (see above). Next, use the check-boxes to turn on the layer that you want to select from and turn off any layers that are above it in the TOC. Click on or near the feature that you're interested in – it will become highlighted in blue once selected.
  - ii. **From within a table.** Open a data table or an attribute table. To select a table record, click the left edge of the row that you're interested in. The selected row will be highlighted in light blue. Use *Shift+click* and *Ctrl+Click* to select or de-select multiple rows.

Note that for attribute tables (i.e. tables containing spatial information so that one feature on the map corresponds to one row of the table), selecting a feature by clicking on the map also selects the corresponding row in the attribute table and vice versa (see image overleaf).

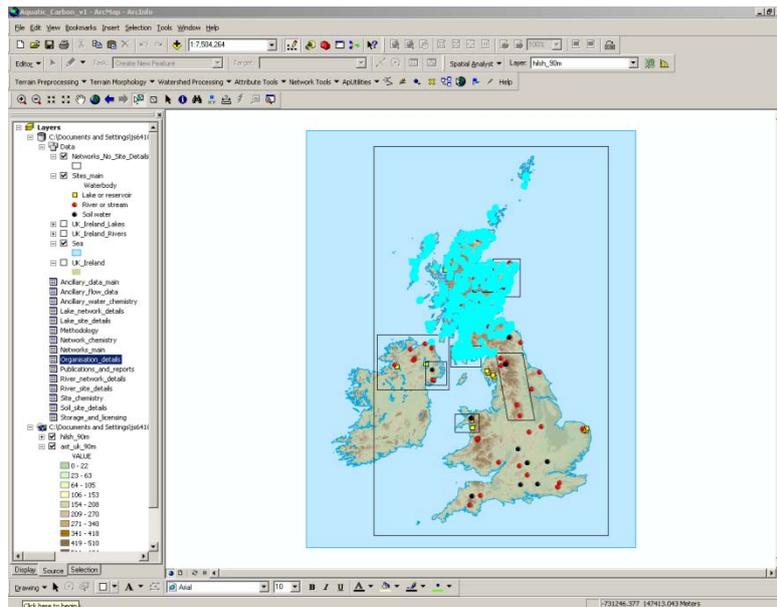
- You also need to be able to *deselect* objects. It is important to **remember to do this** whenever you begin a new query, as otherwise selections from previous work can lead to incorrect results. To clear a selection, go to the *Selection* menu and choose *Clear Selected Features*.



- Data tables can be linked to each other and to attribute tables by using *relationships*. A relationship links a single feature (i.e. a row) in one table with one (or more) feature(s) in another table. Relationships are used to avoid duplication, to simplify the way in which data is stored and to make it easier to get to the information that you need. For example, right-clicking *Organisation\_details* and choosing **Open** will display a simple table containing basic contact information for all of the organisations that have contributed to the database.

The *Organisation\_details* table has two relationships associated with it: *Organisations\_Have\_Networks* and *Organisations\_Have\_Sites*. If you want to see a list of all of the sites associated with, for example, SEPA, try the following:

- First **deselect** any objects that are already selected and then **open** the *Organisation\_details* data table (see above).
- Find “SEPA” in the *Organisation\_name* column and click the left edge of this row to select it. It should become highlighted in blue.
- Click the **Options** button on the *Organisation\_details* data table and choose **Related Tables**. The two relationships mentioned above are listed. Click to follow the relationship entitled *Organisations\_Have\_Sites: To\_Sites\_main*. The *Sites\_main* table will open and all of the records associated with SEPA will be selected in it (i.e. highlighted blue). Note that Non-SEPA records will still be visible, but they’ll be shown in white (i.e. not selected). If you only want to display the SEPA records in the table, click the **Show: Selected** button on the *Sites\_main* table. Those sites which are not selected are temporarily removed from the list of records.
- Close** the *Organisation\_details* and *Sites\_main* tables and zoom to the *full extent* of the map (see above). Because *Sites\_main* is an attribute table containing spatial data, selecting records in this table also selects records on the map, so you should see all of SEPA’s sites highlighted in blue on the map as well (see image overleaf). Remember that you selected these sites (all 518 of them) by initially selecting just a single record in the *Organisation\_details* table and then following a relationship. Navigating from table to table by following relationships is the basic principle underpinning the relational geodatabase.
- Remember to **deselect** the selected records before continuing.



### Part 3: Elements and structure of the geodatabase

To summarise Part 2, the geodatabase can be divided into three main parts:

- i. **Background map layers** (*Sea, UK\_Ireland*) improve the basic appearance of the map but are not involved in the database itself.
- ii. **Data tables** (from *Ancillary\_data\_main* to *Storage\_and\_licensing*) store the core information from the master Excel spreadsheet (split into manageable chunks) but they do not contain any spatial data themselves.
- iii. **Feature classes** (*Networks\_No\_Site\_Details* and *Sites\_main*) which are underpinned by **attribute tables** storing both spatial and non-spatial information. The appearance (colour, size shape etc.) of map elements in these layers is determined by entries in their attribute tables.

Data tables are connected to each other and to attribute tables via **relationships**. Selecting a record (i.e. a row) in one table and following a relationship will select all of the associated records in the second table. Selections in *attribute tables* also result in objects being selected on the map (i.e. in the *feature classes*), so following a relationship which links to an attribute table will create a map showing your records of interest.

The ability to follow relationships from table to table is the key to getting the most from the database. However, due to the complex nature of the data acquired the relational structure is quite involved, so it's useful to have a "map" showing which relationships link to which tables. The schematic diagram on the next page provides this; it's designed to be printed at A3 size and it's not nearly as horrendous as it looks! It's worth spending a little bit of time describing this figure:

- Each shaded rectangle on the "relationship map" represents a table in the database. The title bar for each box gives the name of the table, as it appears in the TOC of the ArcMap document. The main body of each box lists the column headings associated with each table (i.e. what information each table contains) together with the type of information stored in that column ("text", "bool" etc. ; "bool" stands for Boolean and is used to represent answers to yes/no questions from the questionnaire).
- If you're after a particular type of information, looking at the column headings on this map is a good place to start. For example, suppose you wanted to know which river networks recorded information on DIC. The map shows that the data table *River\_network\_details* includes a Boolean (i.e. yes/no) column for DIC, so opening this table is a good first step. By scrolling along to its right-hand end you can immediately see which networks routinely monitor DIC: there are three rows in the table that have "ones" in the DIC column.
- The red and blue lines represent relationships linking tables in the database: red lines are for networks and blue lines are for individual sites. You can only move between tables which have relationships



linking them on the map. To return to the example from Part 2, suppose you wanted a list of all the publications and reports related to monitoring sites maintained by SEPA. One way to do this would be as follows:

- i. First **deselect** any records that are already selected. You're interested in sites maintained by SEPA, so **opening** the *Organisation\_details* table and selecting the row for SEPA is a good place to start (exactly as described above).
- ii. Find the *Organisation\_details* box and the *Publications & reports* box on the relationship "map" (bottom right corner). You will see that these boxes aren't directly connected by a line, so there isn't a relationship linking these tables directly. You're interested in monitoring *sites* (rather than networks), so you need to consider the blue lines as possible routes. The simplest route from your origin table (*Organisation\_details*) to your destination table (*Publications & reports*) is first to follow the blue relationship marked *Organisations Have Sites* to reach the *Sites\_main* attribute table, and then follow the *Sites Have Publications And Reports* relationship to reach your destination.
- iii. With the SEPA row selected (highlighted blue) in the *Organisation\_details* table, click **Options > Related Tables > Organisations Have Sites: To Sites\_main** to follow the first relationship to the *Sites\_main* table. This table will appear as before with all 518 sites belonging to SEPA highlighted in blue.
- iv. In the *Sites\_main* table, immediately choose **Options > Related Tables** again and this time choose to follow the **Sites Have Publications And Reports: To publications\_and\_reports** relationship.
- v. The *Publications\_and\_reports* table will open. Click **Show: Selected** to highlight only those publications associated with the SEPA data.
- vi. **Close** all the tables and **deselect** any selected features before you continue.

Looking at the relationship map, you should see that there are only two feature classes (i.e. layers with attribute tables containing spatial information) in the whole geodatabase: *Networks\_No\_Site\_Details* and *Sites\_main*.

- *Networks\_No\_Site\_Details* consists of polygons defining areas in the UK. These areas relate to networks where respondees chose not to provide individual site details. The data associated with these polygons is therefore more limited than for individual sites. The polygons **only** apply to networks with partial information; those networks with individual site details should be accessed either via the *Sites\_main* layer or the *Networks\_main* data table.
- *Sites\_main* shows the location of every site in the database. Some sites are part of networks and some are not, but they're all shown here in this feature class. The key differentiates between river, lake and soil water sites – see the TOC at the left of the screen for details.

Remember that these two tables are the only ones which link directly to the map display. **You need to make a selection in one of these tables in order to see your records of interest highlighted on the map.** You can follow any number of relationships first, and you can build up very complex queries (see Part 4), but when it comes to viewing your records on the map you need to follow a route back to one of these two attribute tables (usually *Sites\_main*) as a last step.

#### **Part 4: Building queries**

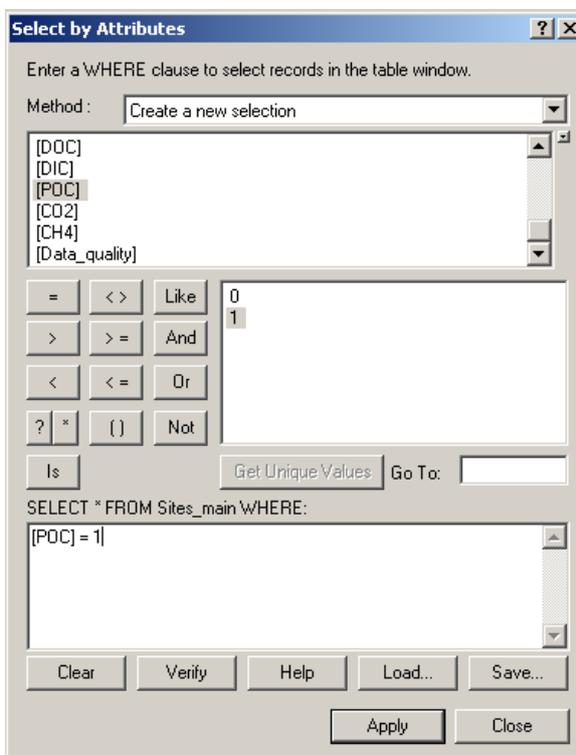
There are many standard ways to build up queries using ArcGIS and a full discussion of them is beyond the scope of these notes. Only the most basic techniques are covered here.

The first step in any geodatabase query is to select some initial records. You can then follow the relationships to select other records which are of interest to you. So far we have introduced two methods for making the initial selection: either by clicking on the map with the *Select Features* tool or by opening a table and selecting some rows (see above). In both of the examples so far, you started off by opening the *Organisation\_details* table and manually selecting the row for SEPA.

Sometimes tables are so large that manually selecting the initial rows becomes impractical. You may also need to build more complex queries based on spatial distribution etc. As a simple example, suppose that you wanted to know which organisations manage sites monitoring POC. Looking at the relationship "map", you will see that the *Sites\_main* attribute table contains a Boolean column for POC, so you could select all of the sites which

have “ones” in this column and then follow the relationship *Organisations Have Sites* to get to the *Organisation\_details* table. Try the following:

- i. **Deselect** any features that are currently selected. **Open** the *Sites\_main* attribute table and **close** any other tables that are open. Select **Show: All** from the bottom of the table to display all of the records. **Zoom** to the *full extent* of the map.
- ii. The *Sites\_main* attribute table contains several hundred records and, as you will see, about 60 of these monitor POC. You could make your initial selection by scrolling down the list and manually selecting every row where the POC column has a “1” in it, but this would be very tedious. Instead, choose **Options > Select By Attributes**.
- iii. A new window appears which enables you to build up SQL queries for making selections in the geodatabase. The SQL syntax is very particular, but also very powerful; please refer to a comprehensive SQL reference for full details. If you’re familiar with SQL you can type your query directly into the space at the bottom of the window (entitled *SELECT \* FROM Sites\_main WHERE*). If you’re not comfortable writing your own SQL commands, the buttons in this window will help you to build up your query with the correct syntax.
- iv. At the top of the window, make sure that the *Method* option is set to **Create a new selection**. Below this, you should see a list of all of the column headings in the *Sites\_main* table, enclosed in square brackets (e.g. [Site\_name]). Scroll towards the bottom of this list and double-click on [POC]. You will see that “[POC]” is added to your SQL query at the bottom of the window.
- v. With [POC] highlighted in the list near the top of the window, click the **Get Unique Values** button. The space above the button will now contain a list of all the unique entries found in the POC column of the *Sites\_main* table, in this case just “0” and “1”.
- vi. Click the “=” button and then double-click on the “1” option above the **Get Unique Values** button. Your SQL query at the bottom of the window should now read: [POC] = 1. You could have typed this directly, but you need to be sure that the syntax is correct.
- vii. Click the **Verify** button to check that ArcMap understands your query. If it is not successfully verified, check that it reads exactly as shown in the image below. Once it is successfully verified, click the **Apply** button to run the query.



- viii. On the *Sites\_main* table, click **Show: Selected** to filter the non-POC records from the table. You should be left with 60 records highlighted in blue. If you minimise the table, you should also be able to see all of the sites that monitor POC highlighted on the map. Remember that this is because *Sites\_main* is an attribute table (not a data table – see above).

- ix. You have just used a query to make your initial selection. The rest of the process is the same as before: we want to follow the *Organisation\_Have\_Sites: To\_organisation\_details* relationship to see which organisations maintain the POC sites. This process should be familiar by now, and you should find that there are four organisations with POC data for individual sites.
- x. Remember to **deselect** the selected features before you continue.

The above example is very simple, but exactly the same principles can be applied to create more complex queries. The following examples show the SQL syntax for a few illustrative searches. They can all be created using the query builder as described above, or by typing the query exactly as shown into the *SELECT \* FROM Table\_name WHERE* box.

- **Example 1:** Select all networks which have national coverage of aquatic carbon data of any type.
  - i. **Deselect** any features that are currently selected and **open** the *Networks\_main* data table.
  - ii. Choose **Options > Select By Attributes**. Build the expression **[Extent] = 'National'**
  - iii. **Verify** and then **apply** the query. 8 records are selected in the *Networks\_main* table.
- **Example 2:** Select all sites (lakes, rivers or soils) where DOC has been measured for more than 5 years. Display the results on the map.
  - i. **Deselect** any features that are currently selected and **close** any open tables.
  - ii. **Open** the *Site\_chemistry* data table and choose **Options > Select By Attributes**. Build the expression **[Chem\_species] = 'DOC' AND [Record\_Length] > 5**
  - iii. **Verify** and then **apply** the query. 379 records are selected in the *Site\_chemistry* table.
  - iv. Choose **Options > Related Tables > Sites\_Have\_Chemistry\_Data: To\_sites\_main**.
  - v. **Close** the open tables. The sites meeting the search criteria are highlighted in blue on the map.
- **Example 3:** Select all river sites where DOC is measured at the same time as POC. Show the results on the map.
  - i. **Deselect** any features that are currently selected and **close** any open tables.
  - ii. **Open** the *Sites\_main* attribute table and choose **Options > Select By Attributes**. Build the expression **[Waterbody] = 'River or stream' AND [DOC] = 1 AND [POC] = 1**
  - iii. **Verify** and then **apply** the query. 43 records are selected in the *Sites\_main* table.
  - iv. **Close** the *Sites\_main* table to see the selected records highlighted on the map.

### **Part 5: Narrowing selections**

Suppose you wanted to see a map of all the sites which monitor DOC and which also have associated nitrogen deposition data. This is a little bit different to what you've done so far, as you need to apply multiple queries to different tables. The process would be as follows:

- i. **Deselect** any currently selected features, **open** the *Sites\_main* attribute table and select those records with **[DOC] = 1**.
- ii. Follow the relationship to the *Ancillary\_data\_main* table which includes information on whether nitrogen deposition data is available. This should be familiar so far.
- iii. Your selection in the *Ancillary\_data\_main* table currently includes all the records for which DOC is measured. You now need to select a sub-set of these records i.e. you need to make a second selection from within the records that are already selected. The criteria for the sub-selection will be that **[N\_dep] = 1**
- iv. In the *Ancillary\_data\_main* table, choose **Options > Select By Attributes** as usual, but when the new window opens this time set the *Method* option to **Select from current selection**.
- v. Build the expression **[N\_dep] = 1**, **verify** and **apply** the query. You have now made a compound query: in the first table you selected all of the sites which measure DOC; in the second table you narrowed this selection to include only those sites which also have associated nitrogen deposition data.
- vi. To see these sites on the map you must follow the relationship back to *Sites\_main*. Click **Options > Related Tables > Sites\_Have\_Ancillary\_Data: To\_sites\_main**, **close** any open tables and you will see the records matching your criteria highlighted on the map.

### **Part 6: Exporting selected records to a text file**

There are many ways to manipulate database records once they're selected. Most of the techniques are beyond the scope of these notes, but it's useful to be able to export selected records to a text file which can then be used with Excel etc. This can be done as follows:

- i. Use any of the methods described above to select your records of interest in a particular table (these records will be highlighted in blue).
- ii. From the table, choose *Options* > *Export*. Check that the *Export* drop-down menu is set to *Selected records* and click the *Browse* button (a yellow folder icon).
- iii. Set *Save as type* to *Text file*.
- iv. Choose the location in which to save your output file and type a file name.
- v. Click *Save*, followed by *OK*, followed by *No*.
- vi. The selected records are written to a .csv text file which can be easily opened in, for example, Excel.

### **Part 7: Conclusion**

You should now be able to:

- Open the database and turn the various layers on and off.
- Select sites by clicking directly on the map.
- Select records by opening data or attribute tables and clicking on table rows.
- Select table rows which meet complex criteria using *Select By Attributes*.
- Navigate between related tables using relationships and the relationship "map".
- Make a compound selection by applying a query to one table, following a relationship to another table and then applying a second query to extract a sub-set of records from the first query.
- Create a map of your selected records by following the relationship(s) back to the nearest *attribute table* (not data table).

Note that there are often several equally valid ways to perform the same search; they should all produce the same results.