

Simple Data Manipulation for Local Historians

Guidelines No. 11

Issue 2: December 2005

These guidelines have been prepared in response to a request from the Local History Committee Computer Group for very basic advice on using **databases and/or spreadsheets**. One of the biggest problems for the historian is to be able to manage and retrieve information with a view to communicating it for a specific purpose.

There are, however, many different brands of data handling software on the market and historians usually have their favourite. This guideline, therefore, has asked members of the Computer group to contribute their knowledge of the value of different data handling software so the reader can judge independently. We must, stress, however, that the opinions expressed are purely personal.

The advice is given in the form of answers to the following questions:-

1. [What is a database?](#)
2. [What is a spreadsheet?](#)
3. [What is the difference in practice between a database and a spreadsheet?](#)
4. [Are databases and spreadsheets compatible and interchangeable?](#)
5. [What are the advantages and disadvantages of the products discussed?](#)
6. [Which websites can I use to learn more about each product?](#)

1. What is a *database*?

A *database* is a systematically arranged collection of information which can automatically be retrieved or manipulated. The material collected is referred to as a *database*. Each piece of data is usually classified and contained in a separate *field*.

There are two major types of *database*:

Fixed format: this is where the fields are set up in advance of data input:

Examples include: *Microsoft's 'Access', Corel's 'Paradox', Lotus' 'Approach', or MySQL.*

Free form: this is where fields can be added or deleted as data is type into the computer

Examples include: *Bekon's 'Idealist'*

2. What is a *spreadsheet*?

A *spreadsheet* is primarily a computer programme which is capable of solving mathematical and accounting problems. It is not a well defined structure that holds data.

Instead of fields, it contains cells. These cells can contain hidden formulae and means that if data is altered then, of course, so is the result anywhere in the spreadsheet where that cell is used in calculation. A *spreadsheet* can, however, act as a *database* inserting text of a limited number of letters into the cells. This requires a great deal of concentration on the part of the user.

Examples are *Microsoft's 'Excel', Lotus' '1-2-3', dbase*

3. What is the difference in practice for the local historian using the *database* or the *spreadsheet*?

<u>Qualities/Type</u>	<u>Database</u>	<u>Spreadsheet</u>
Strengths	<ul style="list-style-type: none"> ➤ The fields are defined to contain particular types of data, dates, numbers, text. ➤ Only the correct data can be put into the fields (data validation) ➤ Data sorting can be achieved through simple queries of the whole database. ➤ Queries can be exported and put into a word processing package. ➤ Powerful search available ➤ Relations between tables allow more precision in the definition of tables (The tables allow a depth to the database rather than only “flat” tables.) ➤ Great scalability, depending on the database. 	<ul style="list-style-type: none"> ➤ Excellent for analysis of historical information which contains mathematical aspect ➤ Graphs can be created simply ➤ Sorts data by <i>column (field)</i> or row (item) ➤ This information can be pasted into a word processing package. ➤ Essentially free form for data entry ➤ Very quick to begin using for data entry
Weaknesses	<ul style="list-style-type: none"> ➤ The database has to be well considered before it is set up. ➤ One is restricted to the fields set up and how they were set up. 	<ul style="list-style-type: none"> ➤ No data entry validation ➤ Searching is limited ➤ No ability to search data between spreadsheets (They are what is called “flat” tables.) ➤ Overcoming the above weaknesses requires programming skills ➤ Limited number of characters per cell

From this simple table, it is clear that the database is the more powerful option. It does, however, require more planning before embarking on the database route. The spreadsheet

is quick and simple – just open a spreadsheet and put in the data – however the spreadsheet does not require any precision of data when it is being entered. The database can enforce it, for instance, a date field will only take data in a date format.

4. Are databases and spreadsheets compatible and inter-changeable?

a. Access to *Excel* and vice-versa

Access tables can be individually exported into *Excel* spreadsheets. Depending on the database's structure, some data could be lost in the export.

Importing *Excel* sheets into Access could lead to corruption of the database, because, unless the data has been rigorously input, fields and cells may not match.

b. *Excel* to *Idealist*

Because of its *free form format* it is very difficult to convert *Idealist* to *Excel* but it can be achieved providing there is uniformity and simplicity in each field. In practice this is unlikely because of the tendency to push on the boundaries of fields losing the regularity of data in the field.

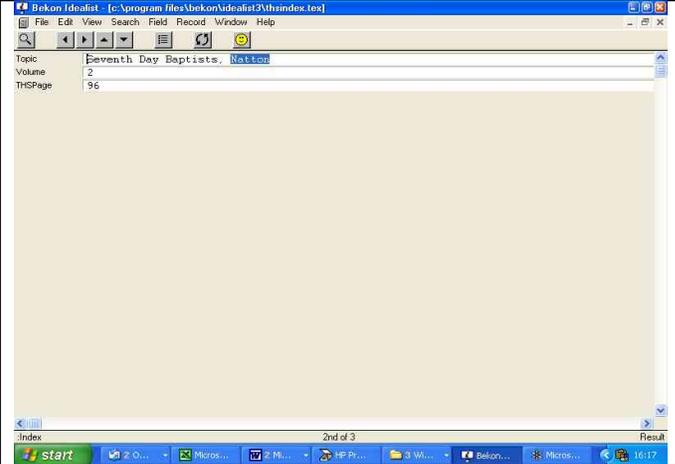
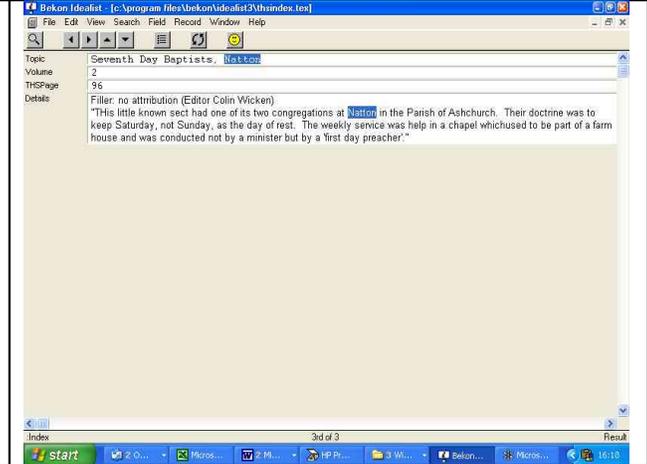
If there is a desire to develop a **database with mathematical aspects** that might require analysis, my advice would be to input data firstly into *Excel* in order to analyse mathematical data. After that exercise, any mathematical figures arrived at, by using hidden formulae, would need to be converted into simple numbers.[#] The spreadsheet should then be saved as a **CSV (Comma Separated File or Comma Delimited File)** and then *imported* into *Idealist* for use as a database[#].

5. What are the advantages and disadvantages of common products?

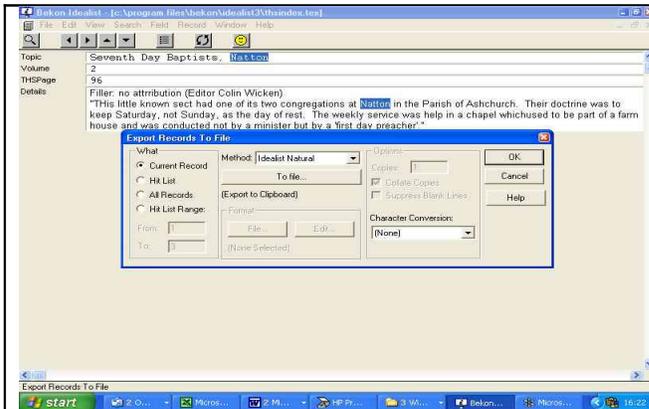
a. *Bekon's Idealist Free Form* database

Advantages

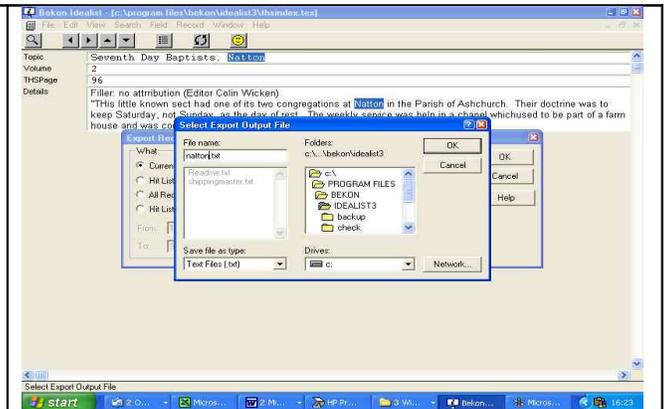
Like any *database*, one usually sets up the fields in advance, but the *versatility* of *Idealist* is this – as data is input and new fields are suggested by the data, these new fields can be added as circumstances dictate. The records themselves are fluid. *Idealist* is text based; the text itself determines its shape, not the fields. In *Idealist* fields allow the text to show itself.

 <p>The screenshot shows the 'Original Record' in the Bekon Idealist software. The window title is 'Bekon Idealist - [c:\program files\bekon\idealist3\thindex.tex]'. The menu bar includes File, Edit, View, Search, Field, Record, Window, and Help. The main area displays the following metadata: Topic: Seventh Day Baptists, Woburn; Volume: 2; THSPage: 96. The record content is mostly blank.</p>	 <p>The screenshot shows the 'Amended Record' in the Bekon Idealist software. The window title is 'Bekon Idealist - [c:\program files\bekon\idealist3\thindex.tex]'. The menu bar is the same as the original record. The main area displays the same metadata as the original record, but with a 'Details' section containing the following text: 'Filler: no attribution (Editor Colin Wicken)'. 'This little known sect had one of its two congregations at Woburn in the Parish of Ashchurch. Their doctrine was to keep Saturday, not Sunday, as the day of rest. The weekly service was held in a chapel which used to be part of a farm house and was conducted not by a minister but by a first day preacher.'</p>
<p><u>Original Record</u></p>	<p><u>Amended Record</u></p>
<p>Note above start “:Index” which is the name of that Database</p>	

Once a database is complete, then the data can be *saved, exported* and/or communicated electronically (e.g. via email) in an abbreviated form as an *Idealist Natural Database*¹. This method maintains the structure of the database but is easy to communicate with a partner who also has *Idealist*. Exporting to a publishing package that can take pre-formatted text is also possible from *Idealist* (and other databases, it must be said) but because *Idealist* is text based, it is a natural for this sort of export.



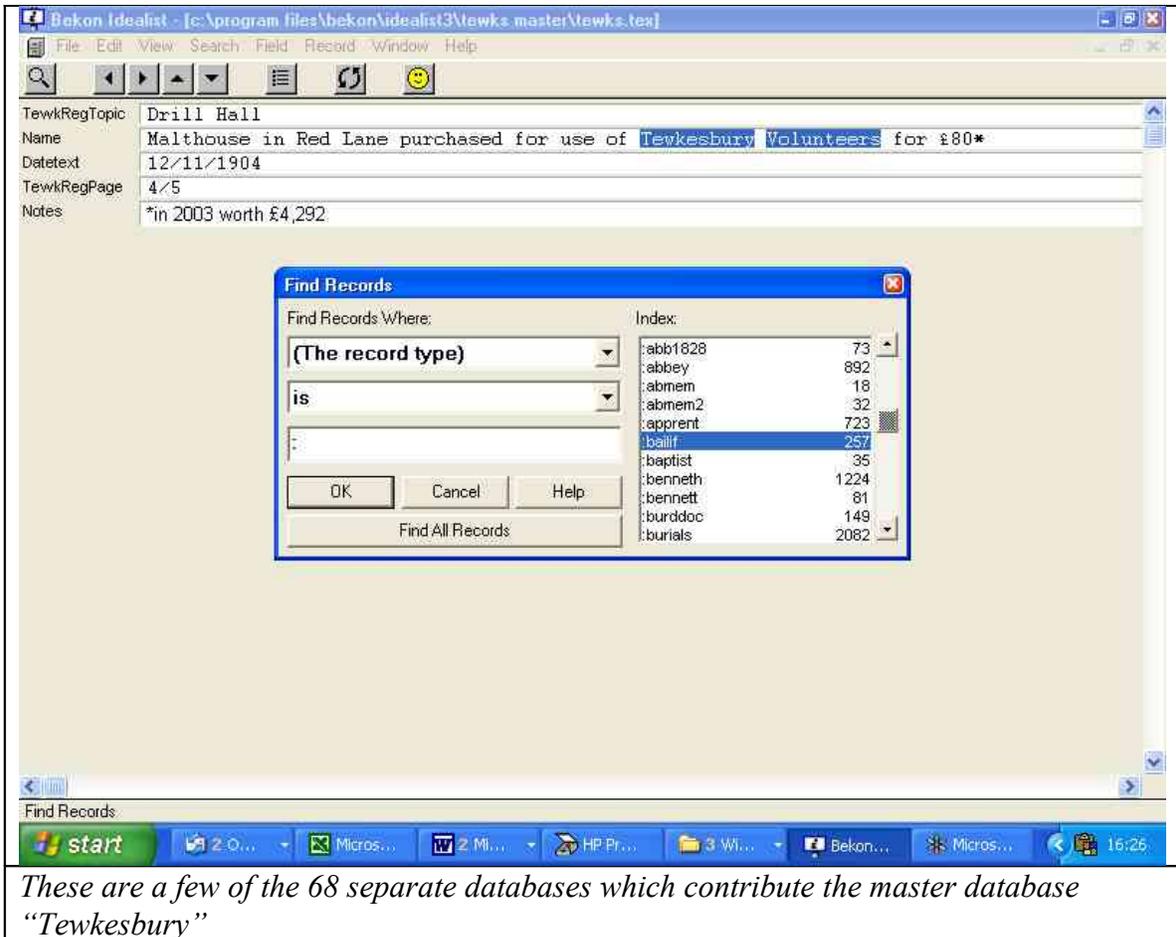
Export Stage 1: note "Method"



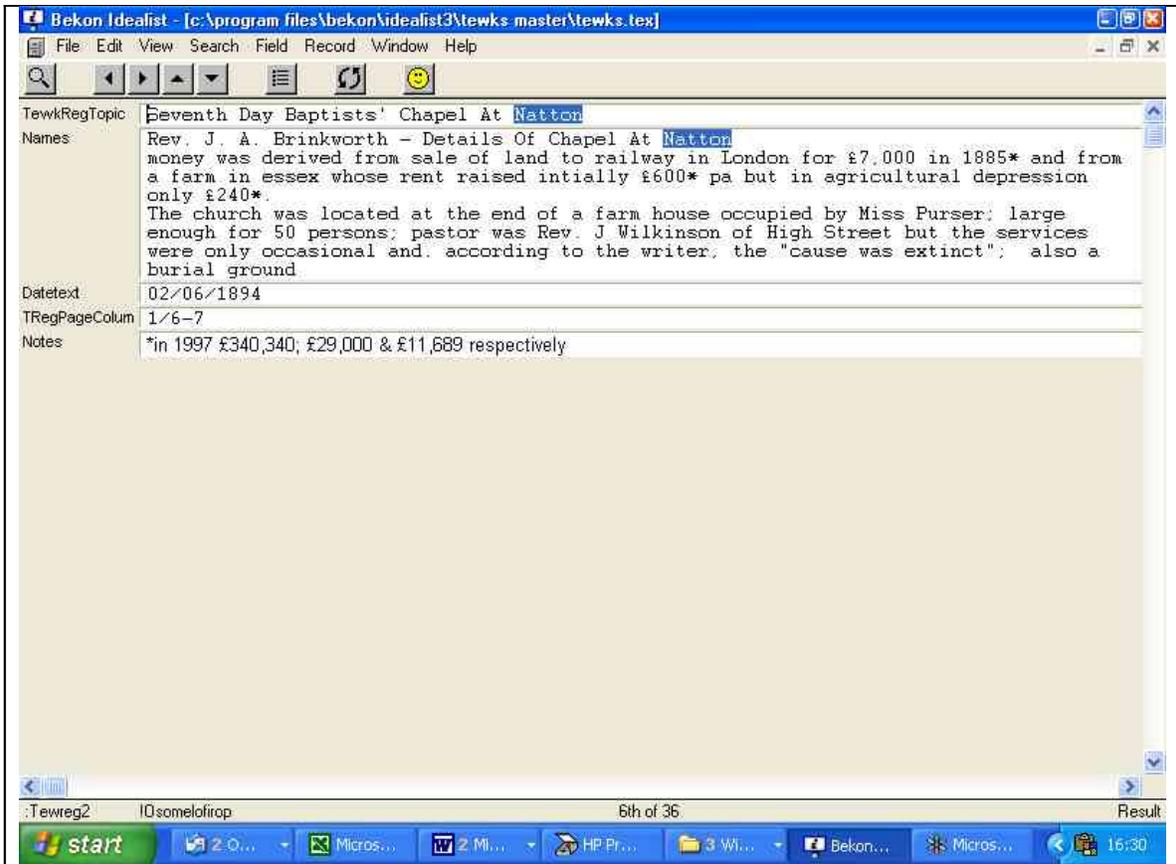
Export Stage 2

¹ It can also be *saved/exported* in *formatted, comma separated* or *tab separated* format.

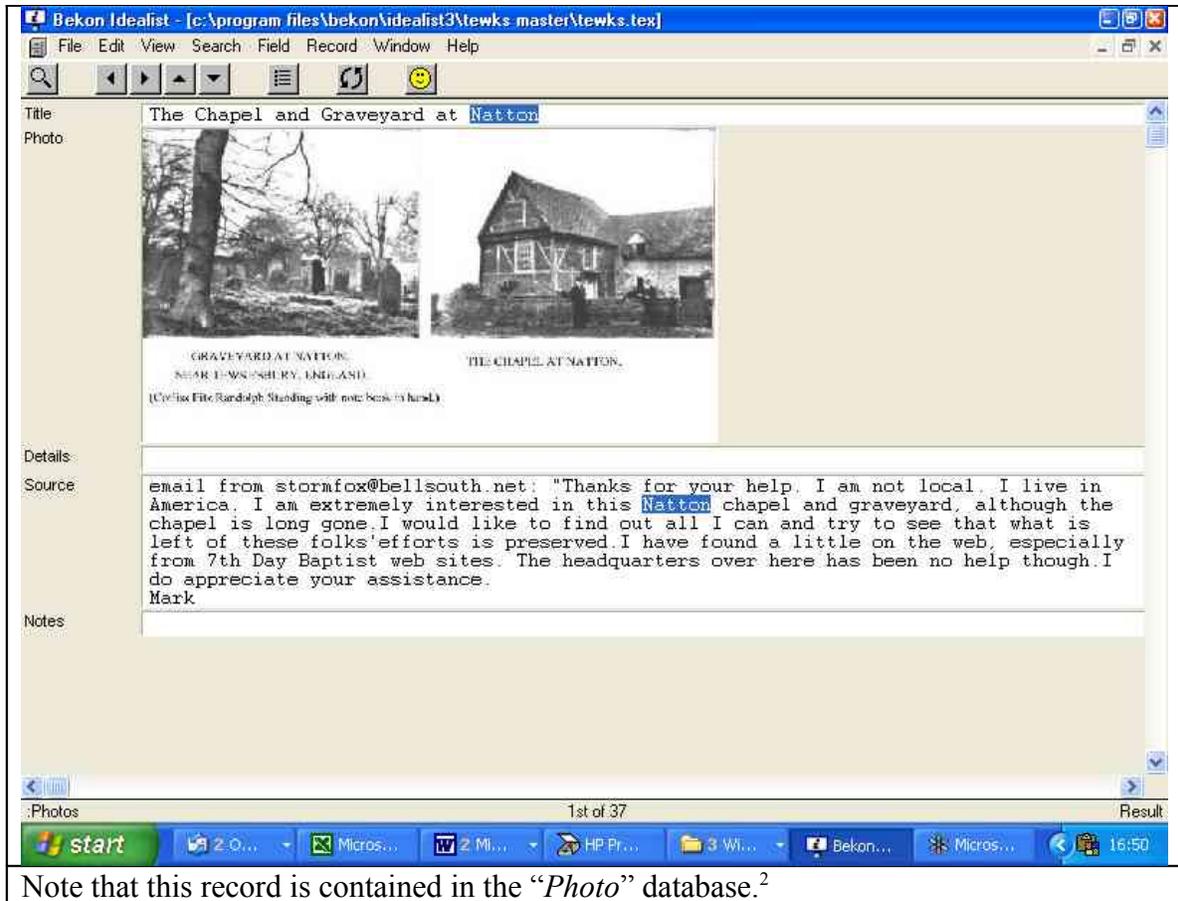
The **most important advantage** is that an unlimited number of databases with different fields and data structure can be imported into a master database using The *Natural* format.



To the uninitiated it looks as though it is now only one database and searches can be made in seconds for one piece of information. The connected information is then highlighted as a *hitlist* for the researcher to peruse.



The Sixth of 36 hits on **Natton**: look above start and you can see that this database is called “:Tewreg2”: this incorporates the data from the index to the local newspaper, the *Tewkesbury Register* inscribed manually once a month by a team of volunteers who started work in 1991. The information is then transcribed by a copy typist.



Note that this record is contained in the “Photo” database.²

This exercise started as a query generated by the website of the Tewkesbury Historical Society: www.ths.freeuk.com that started as a **request** for information from the USA which developed into an invaluable **exchange** of information.

Disadvantages

Not many - but it is not really suitable for doing *complex* mathematical calculations though calculations are possible, as one use of Idealist is as a database for a shop where totals and subtotals are obviously needed.

6. Which websites contain further information about these products?

<i>Bekon</i>	<i>Microsoft</i>	<i>Lotus</i>	<i>MySQL</i>	<i>Paradox</i>
www.bekon.com	www.microsoft.com	www.lotus.com/	www.mysql.com	www.corel.ca/
m	m		/	

John Dixon
John.history@freeuk.com

² Reduced to 400 pixels width via Microsoft's Photo Editor which was part of the package purchased with my Windows XP operating system.