

4

SUPERFILE—a user's view

P. D. Iles*

M. Trueman†

4.1 History of the Lancashire SMR

The Lancashire Sites and Monuments Record was created in 1976, and since 1979 has been housed in the Lancaster University Archaeological Unit, at the University of Lancaster. Initially it comprised a comprehensive map base, together with manual files, and optical co-incidence index cards, and contained about 2000 entries. As the number of entries increased, and computer technology improved it was decided to computerise the Record, using the Central Excavation Unit's software (referred to as Version 1) and a Sirius microcomputer. In 1987, English Heritage undertook the enhancement of County SMR's by supplying a new and more efficient software package, SUPERFILE (written by Southdata). The SMR at Lancaster was transferred to an IBM PC AT running the newly obtained SUPERFILE, and currently comprises c. 11,000 separate records, taking up about eleven and a quarter Mb of storage.

The new software is considerably better and more convenient to use than version 1, in particular in that it allows the use of an on-line database, with dynamic indexes, and the SMR has now been running on SUPERFILE for over two years. In general it has produced very good and satisfactory results; there are, however, some problems with the way that SUPERFILE works, and with how it reacts to the original structure of the Lancashire SMR (first defined in 1976). The system is still used, and fulfills most of our requirements quite well. The criticisms outlined below highlight a number of problem areas, some of which can be avoided by one means or another. These problems are briefly discussed, in order that anyone moving on to SUPERFILE may avoid them from the outset.

4.2 Superfile

SUPERFILE is a suite of programmes, written and compiled in 'C', and comprises a database manager, a sort package, a forms-based access system, a batch editor, and a unit that compresses the data into minimum storage. There is a programme called LOOK, which allows non-forms based work but is rather specialised in its use, and a programme called TAB which produces tabulated reports. As well as these, English Heritage has commissioned programs to load and unload blocks of data. The database manager runs an indexing system, in fact two parallel systems, which considerably speed up access to any specific items of data, this speed of access being a considerable advantage over many comparable systems. The data is stored as a

* Lancaster University Archaeological Unit
University of Lancaster
Physics Building
Bailrigg
Lancaster LA1 4YB

† Lancaster University Archaeological Unit
University of Lancaster
Physics Building
Bailrigg
Lancaster LA1 4YB

a. Input:

```
[ 3008 ][ Bowers House      ][ House:domestic    ][ Post medieval  ]X  
X[ c.1670      ][ Roofed building    ][ SD ][ 66200 ][ 36790 ][ ac ]
```

this is edited with a word processor to remove the sequence

X<RET>X

and superfluous spaces to produce:

b. Output

```
3008 Bowers House House:domestic Post medieval c.1670  
Roofed building SD 66200 36790 ac
```

Figure 4.1: Superfile Forms for Input and Output

tagged ASCII file, and utilizes variable length fields, so that the size of the data field does not have to be declared in advance, nor is any blank space following on from that data stored (contrast this with dBase for example). The forms system (two versions, FORMS1 and FORMS2 exist) is the 'front end' that the user sees almost all the time. It allows form based input and output, query and display of the information in the database. Forms are created or altered by the user, through an editor, and can display or output as much or as little of the data as the user wishes.

4.3 Problems

The problems which we have with the system can be classified into three categories :

4.3.1 Minor faults

The batch loader from version 1 to SUPERFILE leaves spaces in the middle of words in text fields. These spaces have then to be edited out one by one, and record by record. This could probably be written out of the software, as it results from the way that version 1 stores the data, and how the batch loader then reads that data. In practice records are edited as they are needed, rather than spending a great deal of time on editing the entire SMR in one session.

It does not appear to be possible to create forms over 80 characters wide. This limits the ability to output data. For example it is necessary to add tags to a multiple line form, and then to use a word processing package to alter the output to single lines (Fig. 4.1).

The system must be reset and rebooted when the output device is changed (disc or printer). This can be overcome by using Southdata's management program, DIRECTOR (available at extra cost), which allows you to completely unload SUPERFILE without resetting the system.

It is not yet possible to associate onscreen queries with sorts, in order to recover information in a desired order; nor is it possible, as yet, to produce records in a particular order from a list of record numbers entered at the keyboard. Both these features have apparently been commissioned from Southdata but nothing has been seen of them yet.

4.3.2 Medium faults

Altering a record moves it physically in the database and sortfiles are not automatically updated. This can result in records disappearing from any subsequent print which uses a sort, if the sort in question has not been rerun after the editing session. This is a relatively common feature of databases, and is mainly due to the large processing overhead that would be required to update all the sorts for each edit that is made, and the subsequent slow operation of the database.

Linking associated records is not possible with the current Lancashire SMR data structure. Some form of group record number system (eg a segmented primary record number) would be required. This would necessitate a complete rewrite of the data structure for every record.

The manual, although comprehensive, is generally poorly arranged, and unhelpful. It is often necessary to flick from one section to another, and back again in order to find solutions to even simple queries. Whilst the perfect manual (for any software), remains to be written, and several revisions have been made, this is still a bad example of the species.

4.3.3 Major faults

The forms editor is very limited, and awkward to use. It is possible to bypass this editor, and alter or create forms from within your own favourite editor (eg. WORDSTAR or NEWWORD), but it is easy to make serious errors in the form by doing this.

There is no 'online help' available to the user from the keyboard (as is included in dBase III for instance).

SUPERFILE does not allow the use of extended or expanded RAM. In practice all that this means is that there is a finite limit to the size of database that you can process on a DOS based computer (most IBM and IBM clones use a DOS system). We have reached this limit and we can no longer sort the whole of the SMR in particular fields (the TYPE field for example). There are ways around this; for example:

1. Do not undertake very large sorts. Limit sorts to subsets of the data (our current solution).
2. Physically subdivide the database into several smaller entities.
3. Buy a UNIX based operating system, eg. XENIX, (and possibly a new computer to run it) and a UNIX version of SUPERFILE (UNIX does not have the same memory limits as DOS).
4. Use another software package.
5. Alter the data structure to reduce the amount of processing necessary eg. use numeric codes instead of keywords in the TYPE field.

All of the above present their own problems, 1) is not really a long term solution; 2) drastically increases the amount of operator time needed for work over the whole of the SMR; 3) is probably quite expensive, and UNIX is not famous for its ease of operation; 4) would entail learning to operate and tune a completely new system, which would then no longer be compatible with the large number of other SUPERFILE users; 5) would need the complete recasting of all 11,000 records, one by one ...

The most complex problem has been left to last. There are difficulties in the way that SUPERFILE deals with multiple occurrences of sets of fields in a record.

Fig. 4.2 shows a typical multiple entry in version 1 format and the output that would be expected from a very simple version 1 print program. Whilst the SLEX (data

extraction) and SUPERSORT (sorting) operations on the data were time consuming and cumbersome, the end product was a correct set of data. If this data structure is transported straight to SUPERFILE, forms are required that have multiple occurrences of fields written into them (Fig. 4.3). These multiple occurrences of the fields always appear in the order that they were input. The sets of fields within the record cannot be connected or associated together. This may be illustrated by a simple query related to finds of Roman date. Once the appropriate form has been loaded, and the model created (Fig. 4.4a) for this query ie. Roman in the 'General period' box and Finds in the 'form' box, the search can be activated. The record that appears, however, is not what is wanted (Fig. 4.4b). This record does indeed have Finds in the 'Form' box and Roman in the 'General period' box, **but they are not part of the same entry**. The Roman date derives from the entry concerning the Roman road, and the presence of Finds from the entry concerning the Medieval coin. Whilst this may be conceived as a rather trivial example of this problem, it can and does cause a great deal of extra work in editing and re-editing output files to remove such unwanted records. It can also cause a great deal of confusion as to which records are actually required. It also means that it is almost impossible to utilize the TAB section of the SUPERFILE suite.

Again several solutions present themselves :

1. Commission Southdata to adjust the programs.
2. Use another database package.
3. Adapt the SMR so that it no longer uses multiple entries.
4. Concatenate the sets of data that need to be associated, primarily the fields forming the 'Type/General Period/Specific Period/Form' section, into one new field.
5. Put up with it, and continue editing output.

These 'solutions' have their own drawbacks:

1) Seems rather unlikely at present, and may well be impossible to do without great expense and much work from Southdata. 2) again means obtaining, setting up and learning a new system. 3) entails increasing the size of the SMR by a factor of about three, individually editing each record and splitting it up as necessary—this will mean the RAM limit discussed above will need to be confronted. 4) is the English Heritage solution, but has nearly as many drawbacks to it as it solves; for instance you need the data arranged in each possible permutation of the four fields, thus for a record with three entries in this set of fields, you would need a further twelve entries in the record, probably hitting the RAM problem as above. 5) is the current practice, but it is getting to the stage where the extra work involved for the Lancashire SMR may well justify another, more radical solution.

4.4 An easy solution?

Whilst the different solutions in the preceding sections are all possible, the easiest solution to the problems (for us, not for the programmers, or for those who have to pay for it) would be to get Southdata to write patches that a) allowed us to use extended and expanded RAM, b) allowed fields to be associated with each other within a record, and c) allowed records to be associated with each other. This magic program should use the current data and utilize the same commands as before. Whether this is possible is another matter entirely.

a. Portion of Version 1 Data Structure from Lancs SMR

```

35 Primary Record Number
06 District
07 Parish
08 National Grid Reference
11 Site name/main features
13 Type/general period/specific period/form
31 Created - updated by/month/year
33 OS 1:10,000 or 1:10,560 quarter sheet
    
```

b. Sample of SMR entry in version 1 format

```

35 4720
06 Wyre
07 Bleasdale
08 SD 582434
11 Watery Gate/Medieval coins found on site of Roman road
13 Metal:coin/Medieval/-/Finds;
13 Metal:coinhoard/Medieval/-/Finds;
13 Road/Roman/-/Other structure
31 MF/11/1985
33 SD 54 NE
    
```

Figure 4.2:

```

PRN [ 4720] [ SD 54 NE] Parish [ Bleasdale ]
NGR [ SD][ 58200][ 43400][ ] District [ Wyre ]
Site Name [ Watery Gate ]
Site Features [ Medieval coins found on site of Roman road ]
Type [ Metal:coin ] Form [ Finds ]
General Period [ Medieval ] Spec Period [ - ]
Type [ Metal:coinhoard ] Form [ Finds ]
General Period [ Medieval ] Spec Period [ - ]
Type [ Road ] Form [ Other structure ]
General Period [ Roman ] Spec Period [ - ]
Type [ ] Form [ ]
General Period [ ] Spec Period [ ]
Type [ ] Form [ ]
General Period [ ] Spec Period [ ]
Type [ ] Form [ ]
General Period [ ] Spec Period [ ]
Type [ ] Form [ ]
General Period [ ] Spec Period [ ]
Updated - Created [ MF ][ 11][ 1985]
    
```

Figure 4.3:

a. Form to Search Database

```

PRN [      ] [      ] Parish [      ]
NGR [ ][      ][      ][      ] District [      ]
Site Name [      ]
Site Features [      ]
Type [      ] Form [ Finds ]
General Period [ Roman ] Spec Period [      ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Updated - Created [ MF ][ 11][ 1985]
    
```

b) Result produced from database

```

PRN [ 4720] [ SD 54 NE] Parish [ Bleasdale ]
NGR [ SD][ 58200][ 43400][      ] District [ Wyre ]
Site Name [ Watery Gate ]
Site Features [ Medieval coins found on site of Roman road ]
Type [ Metal:coin ] Form [ Finds ]
General Period [ Medieval ] Spec Period [ - ]
Type [ Metal:coinhoard ] Form [ Finds ]
General Period [ Medieval ] Spec Period [ - ]
Type [ Road ] Form [ Other structure ]
General Period [ Roman ] Spec Period [ - ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Type [      ] Form [      ]
General Period [      ] Spec Period [      ]
Updated - Created [ MF ][ 11][ 1985]
    
```

Figure 4.4:

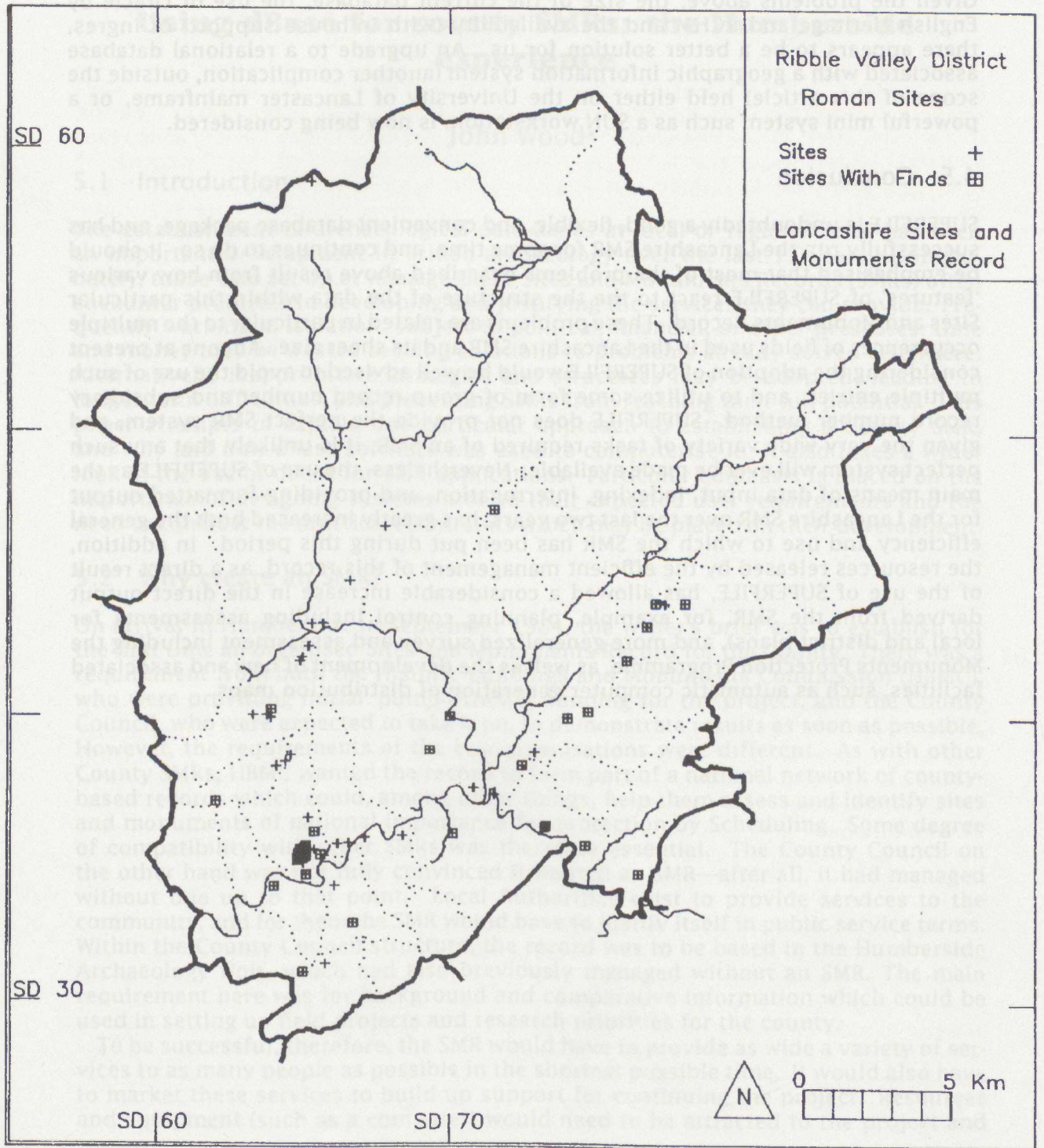


Figure 4.5: Example plot

4.4.1 Current thinking on other solutions

Given the problems above, the size of the current database, the use of Oracle by English Heritage and RCHM and the availability (with in-house support) of Ingres, there appears to be a better solution for us. An upgrade to a relational database associated with a geographic information system (another complication, outside the scope of this article) held either on the University of Lancaster mainframe, or a powerful mini system such as a SUN workstation, is now being considered.

4.5 Conclusion

SUPERFILE is undoubtedly a good, flexible, and convenient database package, and has successfully run the Lancashire SMR for some time, and continues to do so. It should be emphasised that most of the problems described above result from how various 'features' of SUPERFILE react to the the structure of the data within this particular Sites and Monuments Record. These problems are related in particular to the multiple occurrences of fields used in the Lancashire SMR and its sheer size. Anyone at present considering the adoption of SUPERFILE would be well advised to avoid the use of such multiple entries, and to utilize some form of group record number and subsidiary record number method. SUPERFILE does not provide the perfect SMR system, and given the very wide variety of tasks required of an SMR, it is unlikely that any such perfect system will ever be made available. Nevertheless, the use of SUPERFILE as the main means of data input, indexing, interrogation, and providing formatted output for the Lancashire SMR over the last two years, has greatly increased both the general efficiency and use to which the SMR has been put during this period. In addition, the resources released by the efficient management of this record, as a direct result of the use of SUPERFILE, has allowed a considerable increase in the direct output derived from the SMR, for example, planning control including assessments for local and district plans), and more generalized survey and assessment including the Monuments Protection Programme, as well as the development of new and associated facilities, such as automatic computer generation of distribution maps.

