Historical research in the age of the computer: an assessment of the present situation

Schurer, K.

HISTORICAL RESEARCH IN THE AGE OF THE COMPUTER: AN ASSESSMENT OF THE PRESENT SITUATION(1)

K.Schurer(+)

Abstract: Computers are a useful research tool that historians have only recently acquired. The advantages of speed and consistancy that computers can offer to analytical study are well known. Yet to what degree is there a potential danger of research becoming hindered by a misuse of technology? If the computer using historian is to avoid problems of inflexibility he should not allow research to be straight-jacketed by either the computer or its software. Lastly, historians should be aware of possible consequencies that the present revolution in information technology may have on future research.

We are currently led to believe that our society is experiencing a technical revolution, a revolution which has embodied within it a transformation of information technology. One feature of this revolution is that computers have become, or at least are becoming part of our everyday lives. As with all revolutions, this eruption of technology is accompanied by a series of trends, some of which may only be of temporary duration. For example, a whole new range of words and phrases, are gradually entering our vocabulary; school-children boast about the latest items of software and electronic engineers enjoy a prestiged position as the leaders of the revolution. Yet how are historians adapting to this new age of information technology?

Historians, not entirely by accident, have often been cast by people outside the discipline as a traditional, conservative sort, not the kind of people to be dramatically influenced by recent technological innovations. Indeed, this impression is no doubt reinforced by the scepticism shown by some 'traditional historians' in their criticisms of the 'new brand of scientific history'.(2) Unfortunately, the casting of history as a discipline far removed from technical innovation and scientific methodologies is not completely ill-founded. Ten years ago a report assessing the impact of computers on the study of social science within British universities concluded that history was last, a long way down the field, behind the other disciplines.(3) The report showed that eighty-five per cent of all social science departments used a computer, yet only 26 per cent of history departments did so. History's nearest rival was politics with 77 per cent of departments using computers. Converting these percentages into 'real terms', this meant that just eight history departments in England indicated in the survey that they had made use of a computer, and, of course, some of this use may have been fairly minimal. Of individuals, only 5 per cent of historians had used a machine, whereas 35 per cent of all social scientists had. Education was the next lowest subject in which 24 per cent of individuals had become involved with computing.(4) No doubt, during the decade since this report the number of historians using computers has markedly increased, yet the number of other computer using social scientists has

(+) Address all communications to: K. Schurer, Cambridge Group for the History of Population and Social Structure, 27 Trumpington Street, Cambridge, CB3 8JJ, England

expanded also. Consequently, if a comparable report were to be conducted now, it might well reveal that the gulf between historians and their fellow social scientists had widened. In short, of technological innovations affecting historical research the computer is to be found a long way down the pecking-order, behind both the development of the printing press and the invention of the photocopier.(5) Even when a computer is used, often it is chosen only for its word-processing facilities.(6)

So far a disparate relationship between the historian and the computer has been described. Obviously this representation is not entirely the case or else periodicals such as this and Historical Methods could not exist. Much cooperation and interchange of ideas and experience does take place between the computer using members of the historical disciplines. One aspect of this atmoshpere of cooperation that historians have been allied to is the setting up of national date archives; to collect, house and distribute machine-readable datasets relevant to research in the social sciences. In England a Data Archive was established at the University of Essex in 1967, funded by the Economic and Social Research Council (ESRC)(7); similar archives exist in other countries.(8) Although the bulk of the historical datasets available from the Essex Data Archive are based on the nineteenth-century census returns (notably those from the national two per cent sample of 1851, deposited by Professor Michael Anderson of Edinburgh University), the collection also includes files of aggregated parish register data, insurance policies and marriage registers.(9) The use of the archive is not restricted to academics, anybody can deposit data with the archive, indeed one local family history society have alreadey done so.(10) More important is the fact that the ESRC requires that any data generated as a result of their funding should be deposited with the archive. However, unfortunately the situation is not as good as this requirement suggests since those depositing data with the Essex Data Archive, whether funded by the ESRC or not, probably form a clear minority of computer-using historians. The computerised datasets from many research projects, particularly post-graduate thesis work, have sometimes been deleted or lie disused in various university departments or computer services throughout the country. It would appear that once having spent so much time and effort in collecting and making machine-readable a series of historical documents, many historians wish to safeguard the information from other people, protecting it from potential critics, regardless of whether or not they have any intention to carry out further research on the data. The same is true of manual work, few of the many transcripts and indexes made of historical documents, which are of potential interest to other academics, family historians and genalogists alike find their way to appropriate local record offices. It seems ironic that although the photocopying of documentary sources can raise considerable problems due to copyright, researchers can make machine-readable or hand transcripts of them without any liability. Perhaps archivists should insist more firmly that copies of any transcriptions must be placed with the appropriate record office and/or other relevant data depositary.(11)

It is easy to simply state that all machine-readable transcripts and data files should be sent to a data archive, yet the practicability of the situation is, unfortunately, not quite so straightforward. In common with most other kinds of archive, data archives are often under-staffed and under-financed, and, consequently, may well not entirely welcome, or be able to cope with, a flood of depositions. Even if archives did have the facilities to handle such an influx, questions concerning potential usage would probably have to be asked. For example, there seems to be a reluctance for academics to re-work 'second-hand' data, and the degree to which local and family historians would want to use computerised data stored magnetical-

ly, rather than appropriate 'hard-copy' print-outs of the information they require is open to question. What good is a library if no-one wants to read or borrow the books? This issue is of particular importance if the chief criterion for the funding of such an archive is the quantity of users rather than the quantity of depositors and quality of data.

The question of data-quality leads indirectly to what is probably the most problematic issue in relation to data interchange and cooperation. That is the question of utility; in simple terms, of how much use will any computerised data files be to a secondary user? This issue is interwoven with the debate concerning standards of data interchange which has been prevalent among the disussions of computer using social scientists over the last decade, and yet still seems to be without any general agreement or forthcoming solution.(12) One fundamental root of this problem is that, as has already been pointed out, historians are not great users of computers, and of the few who do use the computer, some undertake the research with very little understanding of how the computer actually functions. To these users, the computer is just an amorphous box which produces figures for them to analyse. This situation is unfortunately frustrated by the fact that the advice many historians are kindly offered by their computer understanding colleagues in the science departments often shows little respect for the integrity of the historical source material being computerised. It is often assumed that a software package, such as the much used Statistical Package for the Social Sciences (SPSS), will meet all of the historians computer requirements.(13) In terms of the data file manipulation and statistical calculations available from SPSS this assumption may well be true. However, a problem is that the traditional way of preparing information for this type of packaged analysis is to divide the data up into single records of eighty character length, consisting of a limited number of numerically coded variables or fields.(14) In doing this the inherent logic of the historical document is immediately sacrificed to an obsolete computer technology. Thus, many historians are initiated into the world of comupter based research with the approach that if the work cannot be undertaken within SPSS then it cannot be done at all. It is clear that such an approach is unsatisfactory.

Attempts have been, and are being made to remedy the situation and a framework for historical data analysis is being constructed. In general terms, it seems clear that upon computerising an historical document, a distinction needs to be made between data input and data collection. Data should be collected replicating the source as closely as possible, delineating the structure of the document in terms of both logical entities (records) and secondary data attributes (variables).(15) Information should be collected in as full a form as practically possible, rather than in as full a form as the current research project appears to require. This is important since one can never fully anticipate how future research will develop, let alone anticipate the research needs of others who may at some future point wish to use the data. Reserchers all too often retrospectively regret that they did not include an item of information left out of the datafile, yet by then it is too late. For the same reason, the stage of data processing, in which truncation, abbreviation and coding of information occurs, should be separated from the stage of data collection, with all processing being undertaken inside rather than outside the computer.(16)

For documents of a standard format such as census records, taxation lists and parish registers, it is relatively easy to design a format which will represent the logical units of the original document in machine-readable form. A format has already been suggested elsewhere for the enumeration books of the British censuses(17), while similar sources may be collected

along the lines of the formats illustrated in figures 1 and 2. A printed burial register (figure 1) may be recorded in a simple form in which lines tagged 40 indicate the start of a new page, lines tagged 50 relate to the individual entries and lines tagged between 10 and 32 referencing the document.(18) The columns across the page, dividing various pieces of information are indicated by slashes (/), and logical subdivisions within these units, such as between prenames and surnames, are indicated by a comma. Dittoed entries of information are represented by the characters 'DO' and researchers' comments could be added on lines specially tagged for this purpose (80 in this example). The end of the file is recorded explicitly by a line tagged 90. A similar framework can be applied to a document such as a tithe award (figure 2). In this example information on each tithable parcel of land starts with a line tagged 50, on which is recorded the names of the owner and the occupier. Subsequent information on individual portions of land within a particular parcel (fields) are recorded on lines tagged 60. In this case the information given is as follows: the reference to the field on the accompanying tithe map; the name of the field; the usage; the size; the amount payable to the church; and the amount payable to the tithe owner. Information on each land parcel ends with a line tagged 70, which gives totals of the quantity of land and the amounts payable to the church and owners of the tithe. In the case of land parcels which consist of just a single field or holding, no line 70 is recorded and the end is signified by a subsequent line tagged 50. Columns with no information are represented by a single dash (-) unless they are trailing, in which case they are omitted.

The interests of historians, however, are not confined to documents of a standard form and logical consistency; wills, deeds and even parish registers often have little or no coherent logical structure. Faced with this predicament computer using historians have often adapted the data to fit their programmes by juxtaposition, abbreviation and retention of only the information that interests them. Consequently the dataset is rendered near worthless for the purposes of archival use. A more satisfactory approach to this problem would be to collect the data in a free-format fashion replicating the source, adapt the software appropriately and let the computer carry out any adaption, changes or standardisation, retaining at all time a machine-readable copy of the source which can always be referred to or re-worked as necessary. However, unfortunately computers cannot readily understand text typed into them verbatim. In order to produce the sort of analyses historians require the computer has to be given a sense of location. This may be achieved by inserting a series of unique flags or pointers into the text indicating the presence of words or phrases which may be considered by the researcher to form logical entities and attributes (records and variables).(19) For example, an early parish register may be represented in machine-readable form as in figure 3. Baptisms are recorded on lines tagged 50, burials on lines tagged 60, and marriages on lines tagged 70. Attributes relating to these events, such as prenames, surnames and dates are all flagged. All flags start with an asterisk (*) followed by a key symbol with the end of each piece of information being indicated by a slash (/). Alternatively, according to the structure of a particular document, the researcher may wish to interleave lines of free-field and fixed-field formats. For example, in the baptism register of St James, Clerkenwell the incumbent generally made the entries in the format as follows: month; date; prename; 's(on) of' or 'd(aughter) of'; father's prename; parents surname; '&'; mother's prename; 'his wife'; 'born'; date. Therefore lines entered in this standard fashion may be tagged to indicate that they are in this fixed format, whilst lines breaking this convention can be tagged and flagged accordingly as in figure 4. However, although in this example the elements of the standard format are not separated by unique identifiers such as

slashes, just the natural spaces in the text separating the elements, such practice is potentially dangerous since if the recording of a middle name or some other detail went unnoticed, the extra space would invariably throw the computer into disarray.

Unfortunately flagging of data elements is not a straightforward task. If all that is required for retrieval purposes is the flagging of names and occupations then the necessary flags could be quite simple. However, if the researcher wishes to link items of information and retain the context in which the information occurs, then the process of flagging can become extremely complicated. This point is illustrated by work carried out by Dr.Alan Macfarlane and colleagues in a project aimed at making machine-readable every document over the period 1550 to 1750 relating to the Essex village of Earls Colne.(20) Wishing to retain the grammatical syntax of the text, the documents were broken down into a series of entities based round a subject matter. These were then flagged, bracketed and linked to other entities by nesting and numbering of brackets.(21) As figure 5 shows, for a parish register the scheme is relatively easy to implement. However, a lengthy will may require dozens of nested and linked brackets which can only be added after the document has been read, understood, broken down into its composite entities, and the entities linked with each other.

Regardless of the simplicity or complexity of the flags and pointers used by a researcher it is important that a number of points are adhered to. The combination of chararacters used as flags obviously have to be unique and must not occur in the text of the document. Yet equally a flag should not be too long, its meaning should be implicit, perhaps mnemonic, and it must be portable. Therefore, it must not contain characters that cannot be reproduced by other computers, printers or typewriters. Also it must be remembered that flags are only used as pointers to words or phrases that the current researcher feels may be of interest. No set of predetermined pointers will be absolute; other researchers will wish to sub-divide singularly flagged data elements and re-define or re-classify various flagged data elements altogether. Consequently, if data are to be stored in a data archive it is of crucial importance that any flags used do not destroy the integrity of the document and that the original text of the document is always retreivable. It is for these reasons that in the case of free-format documents it is probably best to initially type in the text with appropriate tags indicating the structure of the document, and then add any required flags at a secondary editing stage, thus retaining the flag-free version for reference and archive purposes. Such a policy would also satisfy the demands of the many non-computer orientated historians, (who it must be remembered constitute the majority of historians), who probably only require a legible understandable hard-copy version of the data. Additionally, if those persisting with manual research prepared their transcriptions in line with the conventions used in the first of these stages, then high quality transcriptions could be fed into a computer via an optical character reader, such as a Kurzweil Data Entry Machine (KDEM) and then flagged and formatted as required by computer-using historians.

Many of the points that have been made look slightly to the future, to a situation where data can be referred to, deposited with, and taken out of an archive in much the same way as we currently use libraries. Nothing has been said about the host of technical problems surrounding the accomplishment of this situation; about the problems of incompatability of storage media and storage formats. In the past the development of new technology has adjusted the nature and intensity of many of these problems and will, no doubt, continue to do so in the future. For example, a recent innovation has been

the development of the multi-format floppy disc copying micro-computer, called the 'magic machine' for short. This computer, which is capable of handling over seventy different storage formats, has been installed at the Essex Data Archive to offer a disc transfer service to those who are unable to transfer information between micro-computers at their own institutions. However, again it must be clearly stated that regardless of the technical advances, such a service would be dramatically devalued if the historical data being transferred could not be satisfactorily interpreted as a result of data reduction or subscription in the form of various codes and classifications.

Still looking towards the future, words of caution have to be sounded not only about historical sources being converted into an historical database, but also the data that are being created now for the historian of the future. To what use will the historian put the machine-readable documents that our society is currently creating? Will he be able to use them at all? Already the discs on which the 1960 American census data are stored cannot be read because the disc drives are no longer made and all of the old ones have been scrapped. Similarly the magnetic tapes holding the British 1961 census have already decayed to the point that researchers wanting special tabulations from them have been refused for "technical reasons".(22) Also, what of the machine-readable information and data being compiled now for the historian of the future; accounts of small-businesses, wage-books of companies and the records of schools and local government. Will the information the historian wishes to consult actually exist or will the files have been written-over as they become out-of-date and superseded? What will the literary historian make of files containing no drafts, just completed prose, neatly edited with all trace of initial, superceded versions deleted? How will the historians adapt to a society in which communications, personal and business, are sent electronically and only stored under special circumstances? When files have been stored, will the historian actually be able to read these on his computer, or will he have to use some specialised machine to do so? Perhaps the historians and archivists of the future will need to add the study of obsolete electrical engeneering technology alongside palaeography in order to read the 'documents' of past societies.

NOTES

1 An earlier version of this paper was given at a conference held by the Society of Genealogists in London; the text of which subsequently published in Computers in Genealogy.

2 Cf. the discussion between the two schools of thought in R.W. Fogel and G.R. Elton, Which road to the Past? Two Views of History, New Haven 1983

3 Computing and the Social Sciences: A Report to the SSRC, Social Science Research Council, London 1973. See especially Appendix II, pp. 23-25. See also R.Schofield, 'English historians and the computer', Historical Methods Newsletter, 7, 1974, pp. 111-114.

4 One must not, however, discount historical research that may have been carried out under the banner of another subject. Geography for example was high on the list of those possessing and teaching computer skills.

5 See H.J.Hanham, 'Clio's Weapons', Daedalus, Spring 1971, pp. 509-519.

6 Cf. R.McCaa, 'Microcomputer Software Designs for Historians', Historical Methods, 17(2), 1984, pp. 68-74 & R.Jensen, 'The Microcomputer Revolution for Historians', Journal of Interdisciplinary History, 14(1), 1983, pp. 91-112.

7 Until 1984 the ESRC was previously known as the Social Science Research Council (SSRC).

8 For example, the Danish Data Archive, the Steinmetz Archief (Amsterdam), the Belgian Archieve for Social Sciences, the Norwegian Social Science Data Services and, of course, the Zentralarchiv (Cologne) which overseers the publication of this journal.

9 Details of the datasets held are available from the catalogue published by the Archive (University of Essex, Colchester, Essex, England) and from the Archive Bulletin published tri-annually.

10 The West Surrey Family History Society have deposited 1861 census data for Southwark, (Christchurch and St Saviours).

11 B.Colins, 'The computer as a research tool', Journal of the Society of Archivist, 7(1), 1982, pp. 6-12.

12 An interdisciplinary workshop on the creation, linkage and usage of large-scale interdisciplinary sourcebanks in the historical disciplines was held recently to discuss such problems. Examples of some of the difficulties associated with data interchange are given in the papers of this conference (Göttingen, 1985). See also M.Thaller 'First considerations on an International Wokshop on the Creation, Connection and Usage of large-scale Interdisciplinary Source Banks in the Historical Disciplines', (unpublished paper), Max Planck Institut für Geschichte, Göttingen 1985. In England the debate has been joined by computer-using genealogists and often fills the pages of Computers in Genealogy (see especially, Vol. 1(1), 1982 and Vol. 1(7), 1984).

13 N.H.Nie et al., SPSS: statistical package for the social sciences, New York, 1975. Anon, SPSS-x User's Guide, Chicago, Il: SPSS Inc., 1983.

14 The main text book on the subject, although now very much out of date, recommends the use of numeric coding and restricting the record length to 80 columns. See E.Shorter, The historian and the computer, New Jersey 1971.

15 See, for example, R.Schofield and R.Davies, 'Towards a flexible data input and record management system', Historical Methods Newsletter, 7, 1974, pp. 115-124.

16 R.Floud, An Introduction to Quantitative Methods for Historians, (2nd ed.), London 1979, See especially pp. 202-210.

17 K.Schurer, 'Methodology: recording data from original sources', Historical Social Sciences Newsletter, 2, 1984, pp. 8-11 and K.Schurer, 'Census enumerators' returns and the computer', Local Historian, 16(6), 1985, pp. 335-342. The scheme will be discussed in greater detail in K.Schurer and J.Welford, A draft-standard for the collection of post-1841 census enumeration data, (forthcoming).

18 In this example (figure 1) and others given in this paper numeric tags have been used to indicate an idea of levels or hierarchy of information. Researchers can, of course, alternatively use characters or symbols as tags.

19 An example of using such pointers is illustrated in M.Overton, 'Computer analysis of an inconsistent data source: the case of probate inventories', Journal of Historical Geography, 3(4), 1977, pp. 317-326. Although the examples shown in this paper show pointers typed into the text of the document, the information could be structured as required by developing an interactive programme for data entry. See, for example, G.A.Dobberts, 'An on line system for processing loosely structured records', Historical Methods, 15(1), 1982, pp. 16-22.

20 The Earls Colne project is based at the Department of Social Anthropology, University of Cambridge and is described in A.Macfarlane, Reconstituting Historical Communities, Cambridge 1977.

21 C.J.Jardine and A.D.J.Macfarlane, 'Computer input of historical records for multi-source record linkage', in M.W.Flinn (ed.), Proceedings of the 7th International Economic History Congress, Edinburgh 1978, Edinburgh, 1978, pp. 71-78.
22 C.Marsh, 'Computers and Historical Research', unpublished paper given to the Department of Social and Political Sciences, University of Cambridge 1980.

Figure 1 Machine-readable burial register

```
10/DATA COLLECTION FILE/K.SCHURER/100285
20/PARISH REGISTER/CHURCH OF ENGLAND
31/ENGLAND/ESSEX/BRADFIELD
32/1812-1836
40/BRADFIELD/ESSEX/1814
50/17/JOHN,SHIPLEY/BRADFIELD/APRIL 17TH/28/HY THOMPSON VICAR
50/18/-,BLYTH/MANNINGTREE/APRIL 24TH/4 YE/HY THOMPSON
50/19/JOHN,SEABORN/BRADFIELD/APRIL 30TH/63/DO
50/20/S,FOX/DO/JULY 1ST/76/DO
50/21/MARY,TURNER/DO/JULY 6TH/69/DO
50/22/MARY,MORGAN/DO/AUGUST 24TH/INFT/DO
50/23/S,GOYMER/DO/SEPTR 11TH/39/DO
.
.
.
80/COMMENTS.....
.
.
.
90
```

Figure 2 Machine-readable tithe award

```
10/DATA COLLECTION FILE/K.SCHURER/120285
20/TITHE COMMUTATION SURVEY
31/ENGLAND/ESSEX/STEEPLE
32/1839
40/ESSEX RECORD OFFICE D/CT 333
50/ISAAC,BROWN/JOHN,RADLEY
60/12/KINGS 6A/ARABLE/6,2,22/-/1,15,10
60/13/ROADFIELD/DO/6,2,9/-/1,17,-
60/14/6 1"/"2A/DO/6,3,14/-/1,18,6
60/15/4 A/DO/3,3,34/-/1,3,1
60/16/3 A/DO/3,3,37/-/1,3,5
60/17/4 A/DO/4,-,28/-/1,1,6
60/18/BARNFIELD/DO/3,3,5/-/1,1,6
60/19/GARDENFILED/DO/3,1,16/-/-,18,3
60/20/HOMESTEAD/DO/-,2,3
60/21/2 A/GRASS/2,3,25/-/-,3,-
60/22/6 A/ARABLE/6,1,4/-/1,13,2
60/23/5A MARSH/DO/5,1,37/-/1,6,9
60/24/3A MARSH/GRASS/4,-,38/-/-,5,-
70/58,2,32/-/14,7,-
50/WILLIAM,BLAKE/GEORGE,PATTISON
60/202/LOST FIELD/ARABLE/13,2,15/1,6,-/3,-,3
50/,SOCIETY OF QUAKERS/,SOCIETY OF QUAKERS
60/218/BURIAL GROUND/GRASS/-,-,31
.
.
.
.
90
```

Note that the slash in 6 1/2A (no 14) has been surrounded by quotation marks to distinquish it from a slash separating variables. Dashes (-) in the amounts payable and sizes of fields are permissable since these cannot be confused with dashes indicating missing information as these are always recorded between two slashes.

Figure 3 Machine-readable parish register

```
.
.
.

40/1738
50/*P WILLIAM / THEN SON OF *FN SAMUEL / AND *MN MARTHA / *S CARRINGTON / +
  WAS BAPTIZED THE *D 4TH OF SEPTEMBER / 1738
50/*P SARAH / AND *P MARY / DAUGHTERS OF *FN THOMAS / AND *MN BRIDGET / +
  *S BRASTREE / WERE BAPTIZED THE *D 8TH OF OCTOBER /
60/*P SARAH / AND *P MARY / *S BRASSTREE / WERE BURIED THE *D 16TH OF +
  OCTOBER / THE AFFIDAVIT FOR BURYING IN WOLLEN REGISTERD
50/*P ELIZABETH / THE DAUGHTER OF *F JOHN / AND *MN SUSANNA / *S ROWLAND / +
  WAS BAPTIZED THE *D 22ND OF OCTOBER / 1738
50/*P JOSHUA / THE SON OF *FN JOHN / AND *MN ANNE / *S NUN / WAS BAPTIZED +
  THE *D  5TH OF NOVEMBER /  1738
60/*P  SARAH / *S  YELL /  WAS BURIED THE *D  14TH OF NOVEMBER /  A R
60/*P  SUSANNA / *S  PLUMMER /  WAS BURIED THE *D  15TH OF NOVEMBER /  A R
60/*GP  THOMAS / *GS  GROSS /  AND *BP  ELIZABETH / *BS  ING /  WERE MARRIED THE +
  *D  16TH OF NOVEMBER /  THIER BANNS BEING FIRST THRICE DULY PUBLISHED
60/*P  ALICE / *S  KING /  WAS BURIED THE *D  17TH OF DECEMBER /  A REGIST +
  BEING BROUGHT WITHIN EIGHT DAYS THE TIME PRESCRIBED BY THE ACT OF +
  PARLIAMENT MADE FOR BURYING IN WOLLEN
.
.
.
90
```

Note that lines longer than eighty characters are continued onto a subsequent line. The continuation being indicated by a plus sign (+) at the end of the line and by indenting the continuation line by two spaces. In this example the meaning of the flags differ according to the type of event:

Baptisms

*P	=	Persons prename
*S	=	Family surname
*F	=	Father's prename
*M	=	Mother's prename
*D	=	Date of the event

Marriages

*GP	=	Groom's prename
*GS	=	Groom's surname
*BP	=	Bride's prename
*BS	=	Bride's surname

Burials

*P	=	Person's prename
*S	=	Person's surname
*D	=	Date of the event

Figure 4 Mixed free and fixed formats

```
60/JUNE 16 MARY D. OF WILL. COOKE & MARY HIS WIFE, BORN 24 MAY
60/JUNE 19 MARY D. OF JAMES NEWTON & ELIANOR HIS WIFE, BORN 6 MAY
60/JUNE 24 ELIZ. D. OF JOHN STORER & ELIZ. HIS WIFE, BORN 24 MAY
60/JUNE 24 RICH. S. OF ROBT SAVAGE & FRAN. HIS WIFE, BORN 20 MAY
60/JUNE 26 WILL. S. OF JOHN MUNDAY & ANN HIS WIFE, BORN 13 MAY
60/JUNE 26 WILL. S. OF JOHN JENKINSON & ELIZA. HIS WIFE, BORN 18 MAY
60/JUNE 26 SARAH D. OF RICH. SNOW & ANN HIS WIFE, BORN 18 MAY
70/*D JUNE 27 / *P CATH. / D. OF *F WILL. / & *M CATHR / *S SMITH / BUT +
  I BELEEVE THIS IS ILEGITIMATE
60/JUNE 28 WILL. S. OF THO. ORETON & ANN HIS WIFE, BORN 10 MAY
60/JUNE 30 WILL. S. OF WILL. PAGE & SARAH HIS WIFE, BORN 29 MAY
60/JULY 3 JOSEPH S. OF JOHN BARRETT & SARAH HIS WIFE, BORN 20 JUNE
60/JULY 3 THOMAS S. OF ISAAC SUFFOLKE & ELIZA. HIS WIFW, BORN 29 JUNE
60/JULY 3 WILL. S. OF WILL. DRAKE & CATHR. HIS WIFE, BORN 23 JUNE
60/JULY 5 SARAH D. OF DANIEL KENEDAY & JONE HIS WIFE, BORN 5 JUNE
70/*D JULY 15 / *P THO. / S. OF MR *F DEUEL / *S PEAD / GENT. & +
  *M SARAH / HIS WIFE, BORN *B 4 JUNE
60/JULY 17 SAMLL. S. OF JOSHUA ATKINSON & FRAN. HIS WIFE, BORN 14 JUNE
60/JULY 17 DOROTHEA D. OF ROBT NOBLE & SARAH HIS WIFE, BORN 5 JUNE
.
.
.
90
```

Example taken from R. Hovenden (ed), A True Register of all the Christeninges, Mariages and Burialles in the Parishe if St James, Clarkenwell, From the Yeare of our Lorde God 1551, Vol 1, Christenings 1551 to 1700, The Haleian Society, (Registers, volume IX), London, 1884, p.376 .

*D = Date of the event
*P = Person's Prename
*F = Father's Prename
*M = Mother's Prename
*S = Family Surname
*B = Date of birth

Figure 5 Example of linked flagging

A: Parish Register

JOHN THE SON OF HENRY ABBOTT WAS BAPTISED 5TH MAY 1607

[A person] [who has a name] [and who is involved in a kinship relation] [with another person] [who has a name]. [The first person is involved in an event] [on a date].

(P (N JOHN **)(K** THE SON OF **(P (N** HENRY ABBOTT **)))(E** WAS BAPTISED **(D** 5TH MAY 1607**)))**

B: Will

HENRY ABBOTT AND JONE HIS WIFE DO CLAIM FOR HOLD A TENEMENT IN CHURCH STREET

(P *1 (N HENRY ABBOTT **)) AND (P (N** JONE **)(K (1** HIS **)** WIFE **)**
(H DO CLAIM FOR HOLD **(L** A TENEMENT IN CHURCH STREET **))**

P = Person
N = Name
K = Kinship relation
E = Event
D = Date
H = Landholding
L = Description of land

The number 1 in example B links elements to the first person. Examples taken from C.J.Jardine and A.D.J.Macfalane, 'Computer input of Historical Records for Multi-Source Record Linkage', in M.W.Flinn, (ed), Proceedings of the 7th International Economic History Congress, Edinburgh 1978, 2, Edinburgh, 1978, pp 71-78.