# Creating an Electronic Cardfile Example

## *The Correspondence of Sir Simon Archer*

## Introduction

A manual cardfile consists of a box containing cards on which various pieces of information are stored. It is a well-understood and valuable research tool. This paper explores how a relational database can provide a way of representing such a manual cardfile electronically.

An electronic cardfile shares the advantage with its manual counterpart of storing information about a number of disparate items within a single container and in a logical order. It also provides the ability:

> to order the contents by various criteria, as opposed to the single ordering of the manual cardfile;

> to search speedily for specific information;

> to select a subset of items on the basis of various criteria.

Figure 1 represents a simple design for the cards in a cardfile recording information about individuals. The card contains fields for the entry of a Name, Dates and Description for each individual.



Figure 1

In a relational database the information is also stored in fields. The fields are held together as a record (= card). The records are stored in a table (= box). Figure 2 represents a database table containing three records each composed of the three fields: Name, Dates and Description. All the records within a table will have the same number of fields, although all the pieces of information may not be available for each record – just as each card in a well-structured cardfile will follow the same format to make searching easier.



Figure 2

In a manual cardfile you might employ codes for recording certain information. In a relational database you can also include codes and create another table to record the meaning of each code so that it is held within the same database. It is also

possible to create a table for additional information required by only some records, as you might have second cards in your cardfile for some entries.

As we have seen, the immediate advantage of the relational database over the manual cardfile is the ease with which the records can be searched for specific pieces of information and sorted by various criteria. An electronic cardfile can also be used as a powerful tool within project management. Using database management software an electronic cardfile can be created that provides:

a single repository for the information concerning each document;

a method of storing information relating to the progress of the project;

the ability to format all or part of the contents of the cardfile for publication in electronic or print format;

the automatic generation of indexes and other scholarly apparatus.

It is particularly valuable where a project has to keep track of a large number of separate sources. A number of different database packages are available, which will provide this functionality. However, in order to make the most effective use of the functionality it is essential that you should pay careful attention to how the database is designed. The acronym GIGO (Garbage In, Garbage Out) is particularly relevant in this context. A little thought applied to the design of your database will pay dividends in the long run and avoid the unnecessary duplication of effort.

## *Sample Data*

This description of designing a database uses a small sample of actual data drawn from the correspondence of the Warwickshire antiquary Sir Simon Archer. Although the sample is small, it exhibits typical features of such data. The sample of 5 letters come from two sources, as follows:

### Bodleian Eng Lett b1

fols. 1-2: From Sir Simon Archer to Edward Gwynn, Nov. 1627, concerning the Archer pedigree.

fol. 4: Archer to Gwynn, sending him a copy of part of the manuscript of John Leland's Itinerary, undated.

fol. 18: From William Burton, December 1628, attempting to rearrange planned meeting with Archer and discussing Burton's book.

fol. 21: Burton to Archer, 10th February 1628/9, rearranging date of visit and giving directions on how to reach his house.

### Shakespeare Birthplace Trust DR37/Box87

item 82: From Edward Gwynn to Archer giving his opinion on a copy of Leland's Itinerary sent him by Archer.

## Step 1 – Standardizing the data

In its simplest form the cardfile might record only the archival reference for each document, its date and a brief description.

| Reference | Date | Description |
|---|---|---|
| Bodleian Eng Lett b1, fols. 1-2 | 16271100 | From Sir Simon Archer to Edward Gwynn concerning the Archer pedigree. |
| Bodleian Eng Lett b1, fol. 4 | 00000000 | From Sir Simon Archer to Edward Gwynn sending him a copy of part of the Itinerary of John Leland. |
| Bodleian Eng Lett b1, | 16281200 | From William Burton to Sir Simon Archer |

| | | |
|---|---|---|
| fol. 18 | | attempting to rearrange a planned meeting and discussing Burton's book. |
| Bodleian Eng Lett b1, fol. 21 | 16290210 | From William Burton to Sir Simon Archer rearranging date of visit and giving directions on how to reach his house. |
| Shakespeare Birthplace Trust DR37/Box 87/82 | 16261203 | From Edward Gwynn to Sir Simon Archer giving his opinion on a copy of Leland's Itinerary sent him by Archer. |

In this example the dates have been converted into numerical dates in the format YYYYMMDD with the new year assumed to start on 1st January. A value of 0 for year, month or day indicates that the information is absent. This numerical ordering allows the letters to be sorted in date order.

Provided you were consistent in entering the data, a simple cardfile like this would allow you to order the records (rows):

> by source (Reference);

> by date (Date);

> by sender (Description).

Because the records can be reordered by sorting, it is essential that all information, such as the complete reference, is included in each record.

This simple approach has several disadvantages:

> it is dependent upon the consistent entering of data (if you described a letter as 'To Edward Gwynn from Sir Simon Archer' the ability to order by sender would be lost);

> there is no shorthand identifier provided for each letter;

> you cannot order by recipient;

> letters are assumed to be dated by day, month, year rather than 'summer 1627';

> you cannot differentiate between a date on the original letter and one assigned by the editor (e.g. the undated letter sending Gwynn the copy of Leland may tentatively be dated to Michaelmas term 1626 on the basis of Gwynn's reply in the Shakespeare Birthplace Trust).

## Step 2

Some of the shortcomings of the simple design can be addressed by creating more fields (columns) in our table.

| ID | Reference | Order Date | Date | From | To | Description |
|---|---|---|---|---|---|---|
| 1 | Bodleian Eng Lett b1, fols. 1-2 | 16271100 | Nov. 1627 | Sir Simon Archer | Edward Gwynn | Concerning the Archer pedigree. |
| 2 | Bodleian Eng Lett b1, fol. 4 | 16361100 | [Michaelmas term 1636] | Sir Simon Archer | Edward Gwynn | Sending him a copy of part of the Itinerary of John Leland. |
| 3 | Bodleian Eng Lett b1, fol. 18 | 16281200 | Dec. 1628 | William Burton | Sir Simon Archer | Attempting to rearrange a planned meeting and discussing Burton's book. |

| 4 | Bodleian Eng Lett b1, fol. 21 | 16290210 | 10 Feb. 1629 | William Burton | Sir Simon Archer | Rearranging date of visit and giving directions on how to reach his house. |
| 5 | Shakespeare Birthplace Trust DR37/Box 87/82 | 16261203 | 3 Dec. 1626 | Edward Gwynn | Sir Simon Archer | Giving his opinion on a copy of Leland's Itinerary sent him by Archer. |

This design has some significant advantages over the first.

> Letters can be ordered by sender (From) or recipient (To) and this ordering is not dependent upon the way in which information is added to the Description field.

> Date ordering has been retained by using the Order Date field, but the inclusion of an additional Date field allows dates to be specified in conventional ways as required by the editor.

> The ID field provides a straightforward way to refer uniquely to each letter.

However, it is also clear that the design still has disadvantages.

> Much of the data in the Reference, From and To fields is repeated.

> The repeated data must be consistent, if ordering is to work.

If someone entering the data abbreviated Shakespeare Birthplace Trust to SBT, letters with the SBT reference would appear before those with Shakespeare Birthplace Trust in any list sorted a-z by Reference. Names cause a particular problem in this context, since Lady Anne Clifford and Anne, Lady Clifford are recognised by humans as synonymous but not by database software.

In order to avoid the unnecessary duplication of data, we need to adapt our design to incorporate more tables.

## Step 3

In the third step of the design process we create separate tables to record archival sources and correspondent names. This will avoid unnecessary duplication, avoid the problems of inconsistent data entry and mean that changes (e.g. to an archival reference or in the identification of a correspondent) need be made in only one place.

### Table 1 – Sources

The archival references are entered into a separate table and each is given a distinct identifier (Source ID).

| Source ID | Source |
|---|---|
| S1 | Bodleian Eng Lett b 1 |
| S2 | Shakespeare Birthplace Trust DR37/Box 87 |

### Table 2 – Correspondents

The Correspondents are also entered into a separate table and given distinct identifiers (Correspondent ID).

| Correspondent ID | Correspondent |
|---|---|
| C1 | Sir Simon Archer |

| C2 | Edward Gwynn |
|----|--------------|
| C3 | William Burton |

**Table 3 – The Main Cardfile**

In the Main Cardfile table it is now necessary to divide the reference into a source (SRC) identified by a Source ID and Reference (REF) containing the folio or item number. At the same time the duplicated values in the From and To fields are replaced by the Correspondent IDs.

| ID | SRC | REF | Order Date | Date | From | To | Description |
|----|-----|-----|-----------|------|------|-----|-------------|
| 1 | S1 | fols. 1-2 | 16271100 | Nov. 1627 | C1 | C2 | Concerning the Archer pedigree. |
| 2 | S1 | fol. 4 | 16361100 | [Michaelmas term 1636] | C1 | C2 | Sending him a copy of part of the Itinerary of John Leland. |
| 3 | S1 | fol. 18 | 16281200 | Dec. 1628 | C3 | C2 | Attempting to rearrange a planned meeting and discussing Burton's book. |
| 4 | S1 | fol. 21 | 16290210 | 10 Feb. 1629 | C3 | C1 | Rearranging date of visit and giving directions on how to reach his house. |
| 5 | S2 | 82 | 16261203 | 3 Dec. 1626 | C2 | C3 | Giving his opinion on a copy of Leland's Itinerary sent him by Archer. |

# Further Refinements

The design we have arrived at would allow you to order the cardfile by:

> date
>
> author
>
> recipient
>
> archival reference.

You would also be able to select letters on the same basis.

The design described here could be refined further. By adding additional fields to the Main Cardfile table, you could record other information, such as:

> whether a transcript of a letter existed;
>
> comments relating to the contents of the letter.

Similarly, the Correspondents table could be extended to record comments about individual correspondents.

You might also want to add further tables to the design. The Archer correspondence at the Shakespeare Birthplace Trust is contained in a number of different collections. If you wanted to be able to select records by archive rather than collection, you could add another table to associate a collection with an archive:

| Archive ID | Archive |
|-----------|---------|
| A1 | Bodleian |
| A2 | Shakespeare Birthplace Trust |

A second source of Archer correspondence within the Shakespeare Birthplace Trust could then be added to the Sources table:

| Source ID | Archive ID | Source |
|---|---|---|
| S1 | A1 | Eng Lett b 1 |
| S2 | A2 | DR37/Box 87 |
| S3 | A2 | ER1/101 |

By selecting only those records with an Archive ID of A2, a researcher could then identify all letters that were located at Stratford. This functionality could prove useful, if transcripts had to be checked against originals and a list of letters at each location was required.

## Conclusion

This paper has been based on a simplified set of data where:

each letter has only one author and one recipient;

there is only one copy of each letter;

the place of writing and receipt is not recorded;

there is only one date associated with each letter.

The complications of real data may be accomodated within a relational database by an elaboration of the simple design process described here. An actual example of the design of an electronic cardfile based on real project data is provided by the Francis Bacon Correspondence Project.

The essential requirement is to analyse a representative sample of data at an early stage and to produce a design that meets the requirements of the project. It is possible to introduce refinements to a design at a later stage, but conducting a thorough analysis at the outset will pay dividends as research and editing progresses. The database will provide both a means of keeping track of progress and a basis for the automatic generation of outputs, indexes and other apparatus for your edition.