Francis Bacon Correspondence Project

Design for Electronic Cardfile

Introduction

The following describes and documents the design of an electronic cardfile for the Francis Bacon Correspondence Project. The design was implemented as a relational database, using Microsoft Access 2000. The initial data was extracted from the 952 HTML files prepared by the project's researchers at Birkbeck College.

The letters in the collection present a number of features that are common to this type of source:

there may be one or more authors;

there may be one or more named recipients;

the writing may take place over several days and in different places;

several manuscript witnesses may survive;

several printed editions may exist.

Since the letters cover a timespan of half a century many of the people who occur as correspondents change their titles or official titles during this period.

Initial Design

In the first instance the data extracted from the HTML files was analysed to establish its relationship to the core entity of the letter.
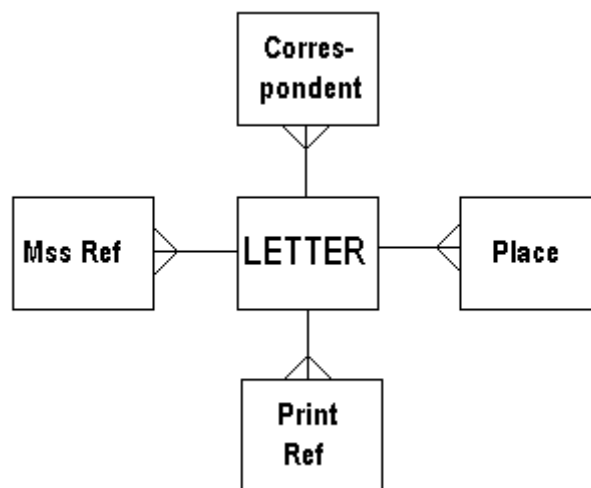

Figure 1: Initial entity-relationship model for Bacon Correspondence data.

Entities/Attributes

	Attribute	*Description*
LETTER	Address	Way in which letter is addressed.
	First Line	First line of letter text
	Summary	Summary of letter contents
	Date	Date of composition

Correspondent	Name	Name of correspondent
	Role	Author / Recipient
Place	Placename	Location of correspondents.
	Direction	From / To
Mss Ref	Reference	Reference to manuscript instance
Print Ref	Reference	Reference to print instance

Each letter had been assigned a number by the Francis Bacon Correspondence Project, which was used as the filename for the HTML files. Since the attributes of the LETTER entity provided no primary key and in order to preserve the link with the source data, it was decided to use this number as a linking attribute (IDENT) for all entities.

As is natural in a sequence of correspondence, many of the correspondents, places of writing and receipt and sources of references were repeated within the data sequence. The entity-relationship model was, therefore, further refined.
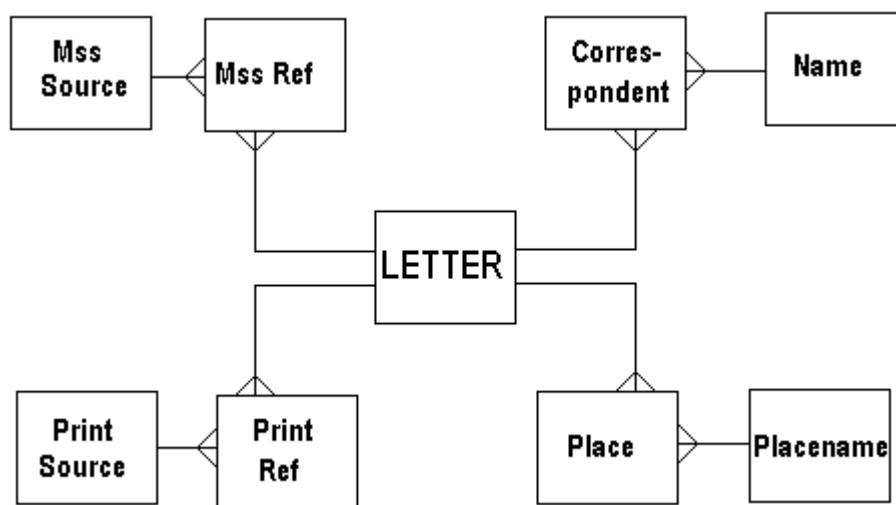
Figure 2: Revised entity-relationship model for Bacon Correspondence data.

Since a handful of letters are recorded in the existing data as having been written from more than one place, I assume that these letters were composed of several sections, to each of which a separate date and summary might be attached. However, since this information is not currently available from the existing data, I have not dealt with the situation in the current stage of the project. In these circumstances it would still be necessary for the LETTER entity to have a primary DATE attribute, which would be used for chronological ordering.

Design Implementation

The design was implemented in an Access 2000 database. During the implementation process the initial design was refined:

to reduce data redundancy;
to facilitate data selection;
to facilitate data extraction.

Access 2000 has two forms of fields for containing text. A Text field can contain up to 255 characters, while a Memo field can contain effectively unlimited amounts of text. The disadvantage of Memo fields is that they cannot be used to sort records. The implementation of the design utilised Memo fields only where the source data for a field exceeded 255 characters.
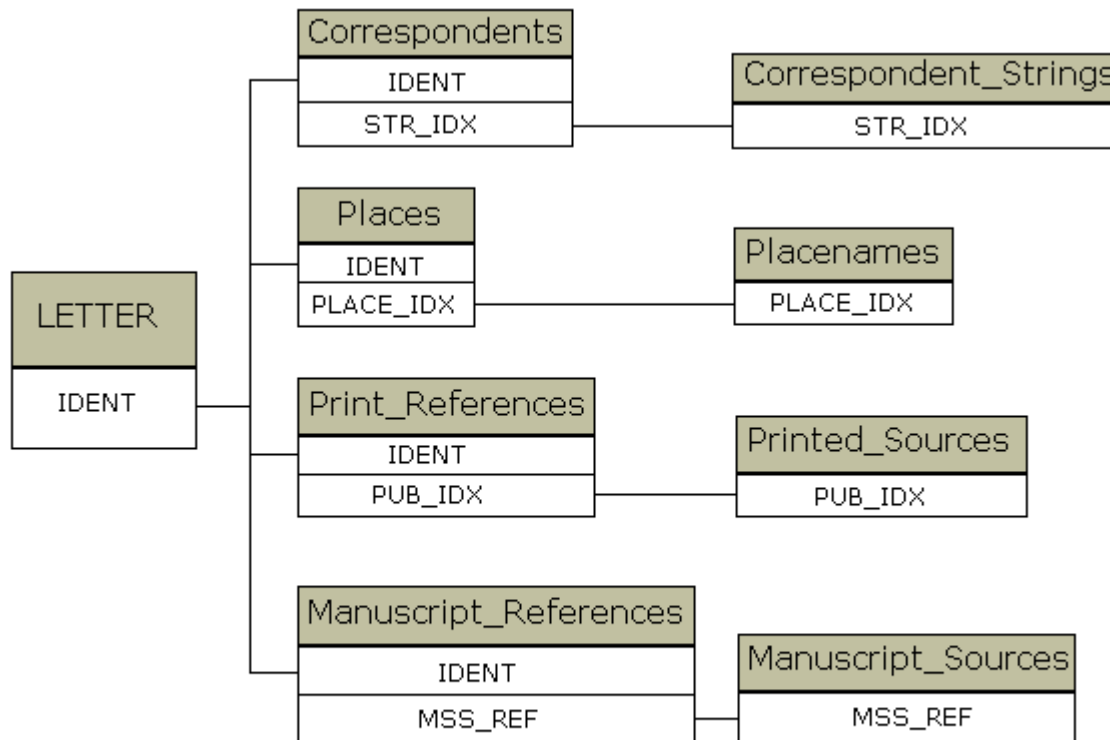
Figure 3: Database design showing field linkages.

Table: LETTER

Field	Type	Comment
IDENT	8 characters	Letter identifier – corresponds to HTML filename
FirstLine	255 characters	First line of letter
Summary	Memo	Summary of letter contents
Date	50 characters	Textual representation of date
Comment	255 characters	Free text field
ADDRESS	Memo	Way in which letter was addressed.
DIRECTION	1 character	I – letter to Bacon O – letter from Bacon S – Bacon acted as scribe U – uncertain ? – not a letter
ORDER_DATE	8 characters	Numerical representation of date in form YYYYMMDD, used to order letters chronologically.

Table: Correspondent_Strings

Field	Type	Comment
STR_IDX	Long Integer	Name identifier
NAME	255 characters	Name of correspondent
PERS_IDX	Long Integer	Identifier of CELL person.

The field PERS_IDX relates the individuals mentioned in the Bacon correspondence to reference database of early modern people that CELL is developing. If the value is 0, then there is no relationship. (A number of letters are from people identified by initials, who have not been identified or from corporate bodies, such as the universities of Oxford and Cambridge.) Where an individual occurs under different names or guises in the Bacon correspondence, a separate entry in the Correspondent_Strings table is created for each name, but they are linked by having the same PERS_IDX value.

Table: Correspondents

Field	Type	Comment
IDENT	8 characters	See **LETTER**
STR_IDX	Long Integer	See **Correspondent_Strings**
Status	1 character	A – author R – recipient
IMPLIED	Yes/No	Identity implied by contents of letter.

Table: Placenames

Field	Type	Comment
PLACE_IDX	Long Integer	Placename identifier
COUNTRY	20 characters	Country – mandatory field
PLACE	20 characters	Town or village
SPECIFIC	50 characters	Specific location within a town etc.

All placenames must specify a country, so that letters can be selected by country where written or received. The SPECIFIC field is used for precise locations within a town or city. For example, Gray's Inn, Bedford House and Coleman Street all occur as specific places within London.

Table: Places

Field	Type	Comment
IDENT	8 characters	See **LETTER**
PLACE_IDX	Long Integer	See **Placenames**
STATUS	1 character	A – Assumed S – Stated Q – Query
To_From	Yes/No	Yes – letter written to place No – letter written from place

Table: Printed_Sources

Field	Type	Comment
PUB_IDX	Byte	Publication identifier
Description	60 characters	Short description of publication
Full_Desc	150 characters	Full description of publication

Table: Print_References

Field	Type	Comment
PUB_IDX	Byte	See **Printed_Sources**
IDENT	8 characters	See **LETTER**
VOLUME	5 characters	Volume of publication
PAGE_REF	50 chars	Page reference

The PAGE_REF may be long, because a few page references include descriptive comments (e.g. mention of facsimiles).

Table: Manuscript_Sources

Field	Type	Comment
MSS_IDX	Byte	Manuscript identifier
Description	60 characters	Description of manuscript

The Description field might have been divided into two fields, one holding the name of the archive and the other the relevant collection. This was not done, as some of the references did not fit happily into this structure and the more general Description was deemed preferable. This decision may be revisited.

Table: Manuscript_References

Field	Type	Comment
MSS_REF	Byte	See **Manuscript_Sources**
IDENT	8 characters	See **LETTER**
Reference	50 characters	Manuscript reference
Holograph	Yes/No	Yes – holograph letter No – not holograph