Where are all the PDBs?: The Creation of Prosopographical Databases for the Ancient and Medieval Worlds

RALPH W. MATHISEN

For decades now, ever since the beginning of the computer revolution in the late 1960s, the possibility of having masses of prosopographical material available in searchable computer databases has been a gleam in the eye of anyone working on the people and populations of the classical and medieval worlds.[1] To cite just one from many examples, a 1989 article in the *Chronicle of Higher Education* suggested that 'computerized databases open up a new range of questions that can be asked that would hitherto have been unthinkable "without 500 monks at hand"'. [2] This statement is particularly applicable to prosopography, for prosopographical information is exceptionally well suited to computer analysis. All the people who have lived during any period of history have associated with them, potentially at least, recurrent categories of information, such as name, sex, religion, marital status, social and economic class, date of birth and death, offices held, and so on.[3] If such information is converted to computer format, it permits the creation of groups of individuals who meet any number of criteria. What prosopographer would not want to be able to select out of a hundred-thousand person population universe all of the senators who ever visited Spain, or all of the Nicene barbarians, or whatever? The possibilities are endless. The following discussion will

[1] See, e.g., G. Althoff, 'Zum Einsatz der elektronischen Datenverarbeitung in der historischen Personenforschung', *Freiburger Universitätsblätter*, 52 (1976), 17–32; J. Barman, R. Barman, W. T. Kirshaw, 'Prosopography by computer: the development of a data base', *Historical Methods Newsletter*, 10 (1977), 102–8; J. Glenisson, 'Prosopographie et informatique', *Informatique et histoire médiévale*, ed. L. Fossier (Rome, 1977), pp. 227–9; Franz Neiske, 'Die Erforschung von Personen und Personengruppen des Mittelalters mit Hilfe der elektronischen Datenverarbeiten', in *Medieval History and Computers*, ed. K. F. Werner (Munich, 1981), pp. 77–110; Lucie Fossier, 'La prosopographie: Les "fantassins de l'histoire" l'honneur', *Le médiéviste et l'ordinateur*, 10 (1983),1–2; Ralph W. Mathisen, 'Medieval prosopography and computers: theoretical and methodological considerations', *Medieval Prosopography*, 9 (1988), 73-128; Neithard Bulst, 'Prosopography and the computer: problems and possibilities', in *History and Computing II*, ed. P. Denley, S. Folgelvik, C. Harvey, (Manchester, 1989), pp. 12–18; *Prosopography and Computer: Contributions of Mediaevalists and Modernists on the Use of Computer in Historical Research*, ed. Koen Goudriaan, Kees Mandemakers, J. Reitsma and P. Stabel (Leuven, 1995). In general, see, D. I. Greenstein, *A Historian's Guide to Computing* (Oxford, 1994); Charles Harvey, Jon Press, *Databases in Historical Research. Theory, Methods and Applications* (New York, 1996); and Sean Townsend, Cressida Chappell, Oscar Struijvél, *AHDS Guides to Good Practice: Digitising History*. (Oxford, 1999): http://hds.essex.ac.uk/g2gp/digitising_ history/.

[2] *Chronicle of Higher Education*, July 12, 1989, p.A7.

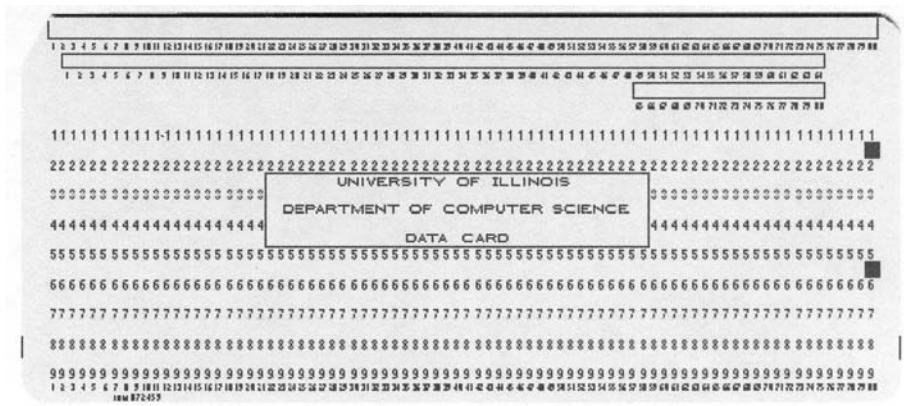
[3] As one prosopographer of the 1970s noted, 'For each element in an historical study the same body of information ... is relevant... One would desire the same types of information (lineage, occupation of parent, family income, educational background, etc.) for each member': J. Lipkin, B. Lipkin Sacks, 'Data base development and analysis for the social historian: the educational status of the beneficed clergy of the diocese of Hereford, 1289 1539', *Computers and the Humanities*, 12 (1978), 11326.

look at how prosopographical databases, 'PDBs' for short, can be created and thus help us to understand how people interacted with each other in the past.

I. An Early Effort

In 1974 I jumped on the bandwagon myself. As a graduate student at the University of Wisconsin and after having spent several years working in the 'real world' as a computer analyst, I decided to create a prosopographical database based on the material in the recently-published volume I of the *Prosopography of the Later Roman Empire* (or *PLRE*).[4] This was in the days of mainframe computers, in this case a Univac 1110, when data were input by means of punch cards. I had 72 columns in which to enter data for each person.

COMPUTER PUNCH CARD



The only way even to begin to do this was to identify categories of information that all individuals shared, such as sex, nationality, occupation, and so on, and assign one-character codes to each of them.

R	Roman (generic)	O	Ostrogoth	Y	Syrian, Phoenician
B	Barbarian (generic)	N	North African	S	Scythian, Steppe Nomad
C	Celtic, British	W	Vandal	J	Jew
U	Burgundian	D	Danubian, Illyrian	A	Arab
L	Gaul	M	Other German	E	Egyptian
K	Frank	G	Greek	P	Persian, Parthian
X	Spaniard	Q	Alan	P	Berber
V	Visigoth	I	Isaurian, Asia Minor	F	Other foreigner
T	Italian	H	Hun		

[4] A. H. M. Jones, J. Morris, J. Martindale, *The Prosopography of the Later Roman Empire. Volume I. AD 260–395* (Cambridge, 1971); see Ralph Mathisen, 'Computer analysis of the *Prosopography of the Later Roman Empire*', *First Annual Byzantine Studies Conference. Cleveland, 24-25 October 1975. Abstracts of Papers* (Cleveland, 1975), pp. 28-9.

CODED VALUES FOR 'NATIONALITY' FROM 1974 DATABASE

Only a few categories, such as names and dates, were written out in full. The result was an outwardly unintelligible set of data images for each person.

DATA RECORDS FROM 1974 DATABASE

FL EUGENIUS	11GC I	GK392N 3AA392023	VC 385 G
EUETERIUS	12 E	LIV 6	LL
EUDOXIUS	12 G	382014	E
EUDOXIUS	12 G	37904C	B
EUCHROTIA	1B B	XI38500B	MAXIMUS L

All the programming for reading, sorting, and outputting the data had to be written from scratch in FORTRAN[5].

EXTRACT FROM FORTRAN PROGRAM FROM 1974 DATABASE

```
READ (5,100) NPARM,IANDOR
 100 FORMAT (I2,I1)
    DO 110 I=1,NPARM
    READ (5,111) IFLD(I),NANDOR(I),IVAL,(NCODE(J,I),J=1,IVAL)
 111 FORMAT (A1,I1,I2,10A1)
    NVAL(I)=IVAL
 110 CONTINUE
  4 LINES=1
    WRITE (6,5)
  5 FORMAT (1H1,6X,"NAME",16X,"SEX",6X,"ORIGIN",4X,"BIRTH CLASS  DEATH
   1 CLASS  MARITAL  REL.  RELIGION   COMMENTS    SOURCE",7X,
   2 "AUTHOR")
  3 READ (5,10) (NAME(I),I=1,4),ISEX,IORIG,IRELIG,IOCC,ILAND,ICBIR,
   1 ICDTH,IAGDTH,IMAR,IRELAT,(IOFF(I),II(I),IDAT(I),IDUR(I),ILOC(I),
   2 I=1,4),(ICOM(I),I=1,9),ISOR,IAUTH
 10 FORMAT (4A6,I1,6A1,A2,A1,A3,4(A1,A1,A3,A2,A1),11A1)
```

It soon became clear, however, that the technology was not yet sufficient to do what I wanted – one could not even enter lower case characters! – and, anyway, I had my dissertation to finish. So work on my own database was suspended for a while.

II. The Promise

Meanwhile, by the mid 1980s, there had been a technological sea change. Improved data entry and data storage technology, combined with the PC revolution and the availability of off-the-shelf database software, opened a Pandora's box of possibilities for

[5] [One of the oldest programming languages, FORTRAN (FORmula TRANslation) was especially suited to numeric computation and scientific computing developed in 1957 by John Backus at IBM; the latest version is Fortran 2003. For more information visit http://www.kcl.ac.uk/kis/support/cit/fortran/f90home.html, accessed 21 July 2006]

the construction of PDBs. Many researchers recognized the value of computerized prosopographical databases. In one of my own grant proposals of that time, I cited nine advantages of a PDB:

1) *Speed of Access*. The larger the database becomes, the greater will be the savings in search time as compared to a manual search.
2) *Accuracy of Access*. The computer itself, to all intents and purposes, will never make mistakes. Most errors will result from mistakes in data entry.
3) *Multiplicity and Diversity of Access*. A computer search of a database can cover many different categories of data at the same time.
4) *Convenience of Access*. A computer search frees one from many hours of tedious work, hours which would be spent more profitably elsewhere.
5) *Ease of Revision*. A hard-copy of a source cannot be revised except by a new edition or by marginalia. A computer database can be infinitely emended. It therefore can serve as a useful, or even necessary, corrective to errors of the past.
6) *Expandability*. A database is 'open-ended': new material can be added from other primary and secondary sources.
7) *Ease of Reporting*. Data from a database can be output in any number of easy-to-read, standardized report formats.
8) *Portability*. A database, or subsets of it, can easily be exchanged with other interested persons in any of several different formats.
9) *Compatibility with Other Biographical and Prosopographical Databases*. The fields of a properly designed biographical database can incorporate individuals with virtually any background, and much of the data from other databases can be merged into it with little revision.

Nowadays, no one would deny the value of having prosopographical information available on computer for sorting, selecting, and analysis.

The future looked bright. Proposals to create shiny new PDBs either from scratch or by computerizing old style card files, using a multitude of different data structures and coding formats, sprouted like mushrooms. These included my own 'Biographical Database for Late Antiquity'– originally dubbed the 'Prosopographical Database for Late Antiquity,' but I soon discovered that anything with the word 'prosopography' in it would never get funded.[6]

So, to what degree, we might ask, has the bright promise for the construction of PDBs been manifested? In an article published in 1988,[7] I discussed twenty projects that were underway to create computerised prosopographical databases.[8] A check on

[6] [Times change. The opposite is true now.]

[7] 'Medieval Prosopography and Computers: Theoretical and Methodological Considerations', *Medieval Prosopography*, 9 (1988), 73-128.

[8] See, e.g., K. Schmid, J. Wollasch, 'Zum Einsatz der EDV in Quellenwerk Societas et Fraternitas', in *Medieval History and Computers*, ed. K. Werner (Munich, 1981), pp. 6976; Robin Fleming, 'A report of the Domesday Book database project', *Medieval Prosopography*, 7:2 (1986), 55-62; Josef Ehmer, 'The Vienna data base on European Family History', *Data Bases in the Humanities and Social Sciences*, vol. 2, ed. R. F. Allen, (Osprey, Fla., 1985), pp. 11316; Caroline Bourlet, 'Construction d'un système declaratif d'aide à l'identifier d'individus dans un corpus prosopographique', *Computers and Medieval Data Processing*, 16:1 (1986), 56.

their current status is revealing: one has been completed (the 'Florentine Catasto' project), one was folded into a later project ('Religious Women', now part of the 'Matrix' project], and two are still in progress (the 'Vienna Family' project, and my own Late Antique database). The rest have either fallen by the wayside or gone underground. Empirically, this suggests that proposed prosopographical databases have an exceptionally high attrition rate. More recent years have seen a continued spate of proposals and 'pilot projects' for PDBs.[9] But, again, wishing is very often quite different from having. When we scan the discipline, we find that we are only beginning to have access to a small number of functional, easily accessible, prosopographical databases for the ancient and medieval worlds. So one question that one might consider is why it has taken so long even to begin to realize the promise of a world of PDBs. Why is it that, after thirty odd years of hope and hype, we do not have a greater number of functional prosopographical databases? Why does it seem that so many projects do not come to fruition? One might begin a consideration of this question by looking at some of the things we have, and have not, learned over the past thirty years.

III. Issues and Answers

During the past thirty years a lot has changed. Whereas general consensus has been reached on several issues that once were hot topics of debate, a few issues still remain under discussion.

1. Speed and Storage

Back in the old days of database creation, limitations on what one could do often were imposed by issues of computer speed and data storage space. The memory of the Univac 1108, for example, was only 65,000 36-bit words. The first PC floppy disks were only 180K bytes. Many will remember those days. But modern PCs, with blinding processing speeds and virtually unlimited hard disk storage space mean that speed and storage no longer restrict us in what we can do and usually no longer are a problem. But, just as a cautionary note, this lack of inhibiting factors still should not be used as excuse for sloppy design.

2. Accessibility

Twenty years ago, PDBs were accessed either by circulation in individual hard copy (on floppy disk), or by contacting the Principal Investigator (PI) and asking for a dataset. Both could be a cumbersome process. Nowadays, the choice *du jour* for accessing a PDB is clearly via the worldwide web, which is increasingly becoming, or already has become, the preferred method for making new computer-based scholarship available quickly.[10] Several PDBs are now available on the web (see below). As data-

[9] E.g., 'The Medieval and Renaissance Italy Prosopographical Database Project' (http://www.slu.edu/departments/history/database.htm); and the 'Religion, dynasty and patronage in Rome, c.440-c.840' database project (http://www.art.man.ac.uk/cla/patronage.htm).

[10] In this regard, one might note the ongoing series of biennial conferences on the topic of 'Ancient studies, new technologies: The World Wide Web and scholarship in Ancient, Byzantine, and Medieval studies': http://www.cisat.jmu.edu/asnt3/.

base to web software becomes more readily available, via programs such as MySQL and dbQwikSite, web-based PDBs will become even more common. Only by this means will PDBs begin to be easily accessible.

3. Hardware and Software Platforms

In the old days, there also was a tremendous amount of debate regarding what kind of computer, or what kind of software, was best for prosopographical applications.These issues have been largely resolved in favour of using any high-powered personal computer, and some kind of off-the-shelf software, most of which is compatible one with another when it comes to exchanging data. An effective prosopographical database should not require the use of any particular kind of hardware or software platform. Indeed, the most effective way to deal with this issue, once again, is to have a menu-driven database accessible via the worldwide web, in which case the only software needed is an internet browser.

4. Nature of Data Reduction

One of the past criticisms of databases was that they imposed too much structure on the information from primary sources,[11] that they created a 'data strait-jacket', and that 'the data basis is an impassable screen between the original sources and historians'. It was suggested that this structure not only created an artificial barrier between the user and the sources, but also was dependent on the interpretive abilities, or lack of same, of those creating the database. True enough, but this, as they say, 'comes with the territory'. In order to be worthy of the designation, any database needs to structure the data from primary sources in two ways. First of all, categories of recurrent information, such a 'sex', 'religion', 'public office', and so on, must be identified. Secondly, if appropriate, recurrent values for these fields, e.g., 'count', 'duke', 'peasant', also must be identified. In order to enforce consistency on data entry, these recurrent values are often replaced by coded values, such as 'MIKN' for 'military: knight', which can be expanded on output.[12] Coded values have a number of advantages:

1. They save space and minimize data-entry time;
2. They can be added as needed and verified on input;
3. They enforce consistency in the data entry: when fields are written out in full, even minor differences (e.g. 'senator' vs. 'senatorial' or even 'Senator') could result in fields being considered to be different by an analysis program; and
4. They provide for the simple creation of subsets and allow for quantitative analysis.

[11] See in particular M. Thaller, 'Methods and techniques of historical computation', in P. Denley, D. Hopkin, eds., *History and Computing* (Manchester, Manchester Univ. Press, 1987), pp. 147-56; Idem, 'The need for a theory of historical computing', in P. Denley, S. Folgevik, and C. Harvey, eds., *History and Computing II* (Manchester, 1989), pp. 2-11; and Idem, *Kleio. A Database System* (St. Katharinen, 1993).

[12] J.-P. Genet, for example, even suggests the creation of such a catalogue if one does not already exist, noting, 'the only reasonable way of handling prosopographical material, is, in fact, to start by writing a dictionary in natural language, and then to extract from such a dictionary a numerically coded database to be statistically studied': J.-P. Genet, 'The PROSOP System', in P. Denley and D. Hopkin, eds. *History and Computing*, (Manchester, 1987), pp. 1918, at p. 192.

Nevertheless, it also has been argued that if a category or value is not included in the data and coding structure, it will be lost to database users.

Creators and users of PDBs have several responses to this criticism:

1. Historians *always* structure their data, whether they are creating a PDB or not. It is up to PDB creators to do a good job in designing their databases;
2. PDB structure and coding are not prescriptive; it only provides a starting point for research. The computer can only do so much. Human intervention is always needed, not only in the course of the creation of a PDB, but especially in the use of a PDB. This includes not only verifying the validity and appropriateness of the data returned, but also judiciously analysing that data. Even when I read an entry in the hard-copy of *PLRE*, I still check the primary source myself. Users of PDBs should do the same.
3. Indeed, the most effective modern PDBs bring the original source documents along with them, either by a pointer to a separate source database or by including the source text within the record, thus ensuring that no source information is ever lost in the creation of a PDB. In light of these considerations, the need to structure and reduce data in PDBs is not now nearly as problematic as it might have seemed in the past.

5. User Transparency[13]

A corollary to the issue of data reduction is that users should be transparent to the technical aspects of database construction, structure, and storage. The user should not need technical programming information in order to be able to access a database.

6. Sorting Out Identities.

PDBs are about individual people. These people must have unique identities within a database. But how do we identify individual people in the sources? If two sources cite people with the same name, how do we know if it is the same person or not? In addition, very often the same person is referred to by different names. Sorting out who's who, either by using a computer algorithm or by human eye-balling, continues to be one of the major problems, if not the major problem, facing the creators of PDBs. Attempting to sort out individuals by computer can take one only so far: as Keats-Rohan has observed, 'Successful linkage can only be achieved in many cases by the historian drawing on personal knowledge and a wide-ranging familiarity with the extensive relevant sources'.[14] In short, sorting out who's who is not a job for amateurs.

7. Database Structure

Another issue that aroused much discussion in the days of yore was the best kind of database structure to use for PDBs, that is, how should the data be stored within the

[13] [Aspect of user friendliness which relieves the user of the need to worry about technical details (like installation, updating, downloading or device drivers).]

[14] K. S. B. Keats-Rohan and David Thornton, 'COEL and the computer: towards a prosopographical key to Anglo-Norman documents', in R.W. Mathisen, ed., 'Late Antiquity and Byzantium,' special issue of *Medieval Prosopography*, 17:1 (1996), p. 244.

database. Some early PDB creators preferred a 'text' (or 'free format') structure in which variable-length database fields were separated by coded delimiters, such as '.na' for name, or '.oc' for 'occupation'.[15]

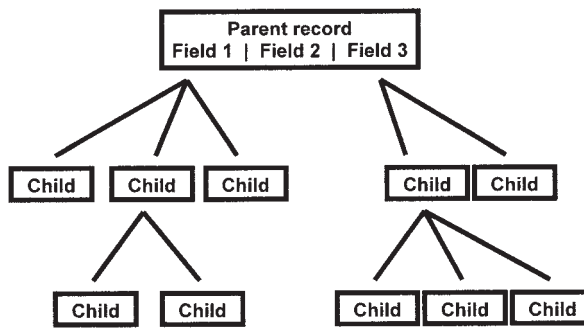
SAMPLE OF A TEXT STRUCTURE

.na Sidonius Apollinaris .bd 432 .bp Lyon .of1 Prefect of Rome .of2 Bishop
.na Eparchius Avitus .bd 410? .bp Clermont .of1 Prefect of Gaul .of2 Emperor

The advantages were that only information that actually appeared in a source need be coded and there was thus a great saving in storage space. The disadvantage was that such databases were very cumbersome to navigate, and could not be used to do quantitative studies without first going through a conversion process.[16]

Others used a 'hierarchical' (or 'tree') structure, in which each field in a primary record had the possibility of having one or more child records.[17]

SAMPLE OF A HIERARCHICAL STRUCTURE



But hierarchical databases were better suited to genealogy than prosopography.

It now is generally agreed that the best structural model for PDBs is the 'relational' model, which was introduced in 1970 by Dr. E. F. Codd of IBM. Codd formulated 12 guidelines that a relational database should implement.[18]

[15] E.g., it 'eliminates the need to code data before input, thus avoiding the practical disadvantages and time-consuming process of coding': Janet Williamson, 'One use of the computer in historical studies: demographic, social and economic history from medieval English manor court rolls', in *Computer Applications to Medieval Studies*, ed. A. Gilmour-Bryson (Kalamazoo, 1984), pp. 51-61; and 'a numerically coded database, though a necessary tool for statistical purposes, is not enough': Genet, 'PROSOP'.

[16] E.g., J. Paul Bischoff, 'Reformatting the Florentine Catasto data for use by standard statistical analysis programs', *Computers and Medieval Data Processing*, 11:2 (1982), pp. 56.

[17] E.g., Gary D. Anderson, 'The SIR Data base management system and its application to census data', in *Data Bases in the Humanities and Social Sciences*, ed. J. Raben, G. Marks (Amsterdam, N.Y., Oxford, 1980), pp. 137–42: 'a case-oriented hierarchical data management system'.

[18] See, e.g., http://www.webopedia.com/TERM/C/Codds_Rules.html

CODD'S RULES FOR RELATIONAL DATABASES (1970)

Rule 1: The Information Rule. All data should be presented in table form.
Rule 2: Guaranteed Access Rule. All data should be accessible without ambiguity.
Rule 3: Systematic Treatment of Null Values. A field can remain empty.
Rule 4: Dynamic On-Line Catalog Based on the Relational Model Rule. Database description accessible in the same way as ordinary data.
Rule 5: Comprehensive Data Sublanguage Rule. The database must be accessible by at least one clearly defined language that supports SQL.
Rule 6: View Updating Rule. Data can be viewed in different logical combinations. Data must be updatable no matter how they are viewed.
Rule 7: High-level Insert, Update, and Delete. Data can be retrieved in sets constructed of data from multiple rows and/or multiple tables. Data can be inserted, updated, or deleted in any manner that they can be retrieved.
Rule 8: Physical Data Independence. User is transparent to the physical method of storing and retrieving information from the database.
Rule 9: Logical Data Independence. How a user views data should not change when the table structure of the database changes.
Rule 10: Integrity Independence. The database language should support constraints on user input that maintain database integrity.
Rule 11: Distribution Independence. User should be unaware of whether or not the database is distributed.
Rule 12: Nonsubversion Rule. There should be no way to modify the database structure other than through the multiple row database language.

A few of these rules are of particular importance for creating prosopographical databases:

1. Data should be stored in a tabular, that is, row and column, format (also known as 'rectangular form' or a 'matrix'), as in an EXCEL file, wherein the 'rows' consist of data records – in our case, representing unique individuals – and the 'columns' consist of different categories of information associated with each record, such as SEX, RELIGION, and so on.

SAMPLE OF A TABULAR STRUCTURE

NAME	IDNO	NAT	LOC	DATE	OCC	OFFICE
Albinus	1	R	Vannes	538-550ca	E	Bishop
Albinus	2	R	Rennes	423-619ca	E	Priest
Albinus	3	R	Angers	567ca	E	Bishop
Baudigyselus	1	R	Angers	567-581ca	E	Bishop
Audoveus	1	R	Angers	581-590ca	E	Bishop
Licinius	1	R	Angers	601ca	E	Bishop
Chaidulfus	1	R	Angers	601-610ca	E	Bishop
Lupus	1	R	Angers	610-627ca	E	Deacon
Lupus	2	R	Angers	627-683ca	E	Bishop
Anonymous	621	R	Nantes	627-683ca	E	Acolyte

The tabular structure makes it easy for users and the computer to locate data fields within each record; and for datafiles to be transferred to different computers, physical storage devices, and software programs without losing their formatting.

2. In addition, each primary data record should be uniquely identifiable, so that a search for a single record will not in fact retrieve two or more records. For PDBs, the unique key usually is created from the person's name combined with a sequence number, for example, 'Albinus 3' or 'Anonymous 621'. Thus, one could look up 'Lupus 2' in a separate 'OFFICES' database to get his offices and in a 'LOCATIONS' database to find all the places he was associated with.[19]

3. Another requirement is that data can be retrieved in different logical combinations, based on values of the fields.

All of the most up-to-date PDBs use some manifestation of the relational database model. The literature for the 'Prosopography of Anglo-Saxon England' (PASE) database, for example, notes, 'Relational databases are now well established as effective tools for historical research, and are particularly appropriate for projects where the recording of the data in a consistently structured way will enable a wide variety of data retrieval and analyses'.[20]

Most PDBs are structured based on a single table that attempts to include as much significant information as possible about a person. The 'single table' model has the advantage of simplicity, but it also has some limitations. In particular, it permits only a single value to be entered per field. Thus, in the format shown here, if 'Albinus 3' is attested to have been a deacon and priest before he was a bishop, there would be no way to indicate that directly in the tabular structure. For 'single table' databases, this problem can be approached in several ways, such as by duplicating a field, that is, having 'OFFICE1, OFFICE2, and OFFICE3'.

TABULAR STRUCTURE WITH ADDED 'OFFICE' FIELDS

NAME	IDNO	NAT	LOC	DATE	OCC	OFFICE1	OFFICE2	OFFICE3
Albinus	1	R	Vannes	538-550ca	E	Bishop	Bishop	Bishop
Albinus	2	R	Rennes	423-619ca	E	Priest	Priest	Priest
Albinus	3	R	Angers	567ca	E	Bishop	Bishop	Bishop
Baudigyselus	1	R	Angers	567-581ca	E	Bishop	Bishop	Bishop
Audoveus	1	R	Angers	581-590ca	E	Bishop	Bishop	Bishop
Licinius	1	R	Angers	601ca	E	Bishop	Bishop	Bishop
Chaidulfus	1	R	Angers	601-610ca	E	Bishop	Bishop	Bishop
Lupus	1	R	Angers	610-627ca	E	Deacon	Deacon	Deacon
Lupus	2	R	Angers	627-683ca	E	Bishop	Bishop	Bishop
Anonymous	621	R	Nantes	627-683ca	E	Acolyte	Acolyte	Acolyte

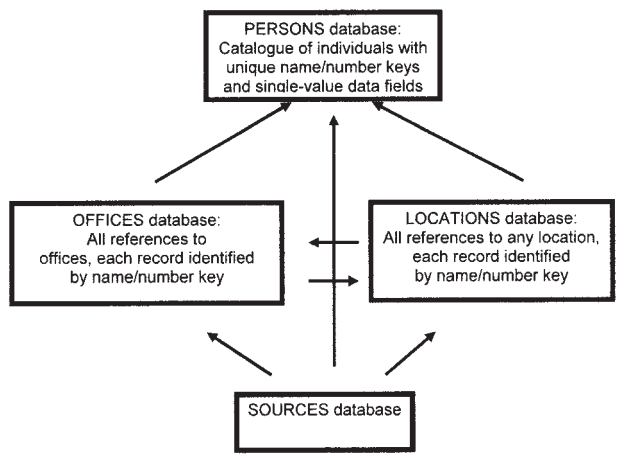
[19] Note that these subsidiary databases are not technically relational because primary items in them (i.e. the values for 'Location') and their associated values (i.e. 'Name') are not unique, although uniqueness could be established by assigning each of the repeated values a numerical sequence number just as duplicated names in the master database are made unique by assigning them sequence numbers.

[20] http://www.kcl.ac.uk/humanities/cch/pase/pase.htm

In this way, only persons having more than three attested offices – probably a very small minority of the population – will be disadvantaged. Another option is to dump information that cannot be accommodated by the existing fields into a catch-all 'BIOGRAPHY' field. But none of these solutions is completely satisfactory.

A rather more complex implementation of the relational structure involves the use of multiple databases dedicated to categories of information that can have multiple values for the same person. Thus, in addition to a PERSONS database that stores singly-occurring information about an individual, such as 'Sex' or 'Place-of-Birth', an OFFICES file would include all of the instances of offices that were ever held by anyone; and a LOCATIONS file could include all instances when a particular location is mentioned. The records in the OFFICES and LOCATIONS databases would be keyed to a particular individual by use of the NAME-AND-NUMBER unique identifying key that all individuals in this kind of PDB must have. In order to create a complete 'biography' for a person, the associated information in the OFFICES and LOCATIONS databases will be retrieved and output to the user. In addition, by giving the offices and locations unique keys, these two databases can be related to each other, all of which will allow for subsets of data based on office and location to be created along with the more traditional subsets of data based on person. In addition, instead of including primary source material in the main PERSONS file, which can lead to repetition and take up a lot of space, one can include an additional SOURCES database, that can be accessed, with the proper source-key, by any of the other databases.

A SIMPLE MULTI-FILE RELATIONAL DATABASE STRUCTURE



As one can imagine, even a simple manifestation of the multi-table relational model can get pretty complicated pretty fast. If followed to its logical conclusion, with a separate datafile for every category of information for which a person could have more than one value, this model could result in having nearly as many stand-alone datafiles as there are data categories.

Finally, a few more words also ought to be said about the word 'relational', because no term has ever been used, or misused, in so many different contexts. For example, in one recent discussion, a writer assumed that the term 'relational' referred to the 'relationship' between the material in the source and that in the database. Other writers suppose that the 'relationship' is between the database and the use made of it by users. And still others suppose that the tabular data storage format is somehow inconsistent with the relational model, when in fact it is the basis for it.[21]

8. Number Of Fields And Tables

According to some schools of thought, a PDB should not have too many categories of information. One handbook, for example, suggests that 'too many tables can make databases daunting and difficult to use'. Others believe, however, that the number of fields and tables ought to be determined by the use to which a PDB is going to be put and by the source material from which it is being drawn. This does not mean, of course, that one should clutter the database with trivial information or information that applies to a vanishingly small number of persons, but that if there is recurrent information in the source material (such as 'CAUSE-OF-DEATH'), that some user might be interested in studying, it had better be included right from the beginning, even if it means increasing the number of data-categories. Some of the most up-to-date PDBs (discussed below) manifest this philosophy.

9. Standardization

In the early days of PDB creation, there was a lot of talk about developing a standardized format for entering and storing prosopographical material so as to facilitate the exchange of information among different databases. Several attempts have been made to implement one format or a 'standard' format.[22] It is doubtful, however, whether any real standardization in data entry formats will ever occur. The data reduction method of any prosopographical database is designed with a view toward the material to be included and the use to which it is to be put. It is advantageous to devise data conversion formats that are suited to particular applications. The use of an all-purpose method would create inefficiencies. Furthermore, the use of a relational database structure in and of itself makes it very easy to transfer data from one database to another, and removes most incompatibility problems. Finally, most of the 'standardized' methods were introduced before the appearance of the current multitude of commercial applications packages, and if anything, these new packages have become the new 'standard' themselves.

[21] E.g., 'Unlike flat-file systems used in the past, the relational model allows the investigator to transcend the document...': http://www.slu.edu/departments/history/database.htm

[22] See, e.g., Manfred Thaller, 'Automation on Parnassus. CLIO – A databank oriented system for historians,' *Historical Social Research*, 15 (1980), pp. 40–65; Hélène Millet, 'Un appel a l'unit', *Le médiéviste et l'ordinateur*, 10 (1983), 1719; Joseff Smets, 'Un logiciel modèle: CLIO', *Le médiéviste et l'ordinateur*, 10 (1983), 811.

IV. Database Types

So much, then, for the theoretical models that can be used to create PDBs. In our quest for the answer to the question of where all the PDBs are, let us now consider the kinds of PDBs that actually have been implemented. Two general types of PDBs have been identified in the literature:

1. 'Restricted' or 'Limited' databases incorporate the persons from one discrete primary or secondary source, such as 'the persons in the Theodosian Code' or 'the persons in the Pisan baptismal records'. Because the data for these databases come from discrete, homogeneous sources, the selection of common categories of information will be largely determined by the nature of the source, and good planning up front will largely obviate the need to change the data structure later on. The users of 'Restricted' databases usually have some potential interest in all the persons in the database. Moreover, this kind of database is well suited for funding proposals, because it actually can be 'completed.'

2. The 'Inclusive' or 'Open-ended' database, on the other hand, proposes to include 'all the people' who lived at some particular time or place. It incorporates material drawn from many heterogeneous sources. This kind of database is much more difficult to design from scratch, as the designers must anticipate both what kinds of information users might want to access and what kinds of information will be provided by the sources from which the database will be constructed. Users of an 'inclusive' database very often are uninterested in large numbers of persons in the database; for example, a user of the *Prosopography of the Byzantine Empire I* who was only interested in persons drawn from John Chrysostom could function perfectly well if all of the other persons were not even there. This kind of database is never 'finished', as there are always new persons to be added to it.

3. A special case is a limited database that has the form of an open-ended database but is constructed from an existing hard-copy prosopographical catalogue, or card-file, in which the limit is imposed not by the source material but by the editors' decisions regarding whom to include in the catalogue.

SUMMARY OF PROSOPOGRAPHICAL DATABASE TYPES

1. Restricted/Limited PDBs
   - Drawn from a discrete, homogeneous source
   - Fields largely determined by source material
   - Users interested in entire database
   - Can be 'completed'

2. Inclusive/Open-Ended PDBs
   - Use a wide variety of heterogeneous sources
   - Fields largely determined by user needs
   - Users interested in subset of database
   - Are never 'completed'

3. Combination PDBs
   - Created from existing prosopographical catalogue
   - 'Limit' imposed by editorial selection process

V. Survey Of Database Projects

Some insight into the question of where all the PDBs are can be obtained by looking at some examples of PDBs that have been created in the past, or are in the process of being created, and consider where they fit into the discipline. However one cuts it, PDBs relate primarily to information about people. There are many sorts of humanistic databases that have people in them, but are not specifically prosopographical in nature,[23] even though with a little data manipulation, some of them could be turned into PDBs. In the following survey of PDB projects, the primary focus will be on databases that are overtly hard-core prosopographic in nature.

A. Biographical Catalogues

Brief mention might be made of onomastic indexes, and what might be deemed 'biographical catalogues', that is, lists of individuals who have various kinds of information assiciated with them but which usually do not have many, or any, search fields except for name. Examples of biographical catalogues include '*De Imperatoribus Romanis*: An Online Encyclopedia of Roman Emperors', a website of Roman and Byzantine imperial biographies,[24] and the several examples of 'saints' websites, which contain biographical material on various categories of saints and ecclesiastics, such as the 'Roman Martyrs Project' at Manchester University,[25] the 'Dumbarton Oaks Hagiography Database';[26] or the 'Women in Medieval Monasticism' project, now incorporated into the 'Monastic Matrix' project.[27]

SAMPLE OUTPUT FROM THE 'ROMAN MARTYRS' PROJECT

ABDON and SENNEN
mm. Romae, sub Decio.—July 30.
BHL 6-8.
AASS. Iul. VII. 137-138.
MH July 29 (402); July 30 (404).
Site: in cem. of Pontianus, on the Via Portuensis (**Pietri**; so too **Amore**, p. 227-8, **MH** July 29). **MH** July 30 adds 'ad ursum pileatum'.
****Abdon and Sennen**
(**BHL** 6-8).
Passio S. Polychronii 2-4, 6, 7-10

[23] E.g., the 'Manchester Late Antique Patronage Project'. In its complex set of tables (displayed below), 'Person' plays only a supporting role: http://www.art.man.ac.uk/cla/patronage.htm
[24] http://www.roman-emperors.org/impindex.htm
[25] http://www.art.man.ac.uk/cla/Romanmartyrs.htm
[26] http://www.doaks.org/Hagio.html.
[27] Suzanne F. Wemple, Mary M. McLaughlin, Heath P. Dillard, Eleanor S. Riemer, 'Barnard College Project on Women's Religious Life, 500–1500 (First Phase: Italy)', *Computers and Medieval Data Processing*, 14.1 (1984), 89; http://monasticmatrix.usc.edu/vitae/

ENTRY FROM 'DUMBARATON OAKS HAGIOGRAPHY DATABASE'

MICHAEL SYNKELLOS
BHG Number: **1296**

Birth Year: **761**	Vita Start Year: **846**
Death Year: **846**	Vita End Year: **876**
Domain: **C/PLE**	Region: **JERUSALEM.MT.OLYMPOS.CONSTANTINOPLE**

Author: **Anonymous**
Edition
M.B. Cunningham
The Life of Michael the Synkellos
(Belfast 1991).
DO call number: TA 17 .B48 (1).
Greek text

Another kind of biographical database can be created from existing biographical catalogues or registers, such as several databases created from lists of Oxford and Cambridge alumni, for which nearly 100,000 records are available, even though not all, it seems, have been incorporated into a single database.[28] The 'Cambridge University Officers' database, for example, is available on the web. It contains 59557 records indexed by name and type of position in the university.[29]

SELECTABLE CATEGORIES IN 'CAMBRIDGE OFFICERS' DATABASE

- Professors
- Readers
- Lectureships: Special Foundations
- Sir Robert Rede's and Barnaby Lecturers
- University Lecturers
- Assist. Dir. of Dev. Studies, and Int. Relations
- Assistant Directors of Research (ADRs)
- Senior Assistants in Research (SARs)
- Demonstrators
- §Assist., Fac., Assoc., & Clinical Lecturers §Instr. in Mod. Languages, and Univ. Teachers Senior Tech., Tech., and Assistant Tech. Officers
- Assistants in Research; and to Professors
- Computer Engineers; Computer Officers

A more ambitious effort to make a biographical catalogue available in computer format was the project to publish the 'Prosopography of the Byzantine Empire' not in hard copy but in computer files. The resultant CD, published in 2001,[30] contains over

[28] A. B. Emden, *A Biographical Register of the University of Oxford A.D. 1501-1540* (Oxford, 1957-1959), Idem, *A Biographical Register of the University of Oxford to A.D. 1500* (Oxford, 1974); T. H. Aston, 'Oxford's medieval Alumni', *Past and Present*, 74 (1977), 340; T. H. Aston, G. D. Duncan, T. A. R. Evans, 'The medieval alumni of the University of Cambridge', *Past and Present*, 86 (1980), 9-86; Ralph Evans, 'The analysis by computer of A. B. Emden's Biographical Registers of the Universities of Oxford and Cambridge', in N. Bulst, J. P. Genet, eds., *Medieval Lives and the Historian, Studies in Medieval Prosopography* (Kalamazoo, 1986) pp. 38194.

[29] See John Dawson, 'ACAD – A Cambridge Alumni Database', *Computers in Genealogy*, 7:8 (2001), 363-75; see also http://www.cus.cam.ac.uk/~jld1/ACAD/); http://www.cus.cam.ac.uk/~jld1/lists/ acohead. Html.

[30] *The Prosopography of the Byzantine Empire I: (641–867)*, ed. J. R. Martindale (Aldershot, 2001). See J. R. Martindale, 'The Prosopography of the Byzantine Empire', in R. W. Mathisen, ed., 'Late Antiquity and Byzantium', special issue of *Medieval Prosopography*, 17:1 (1996), 169-91; Dion C. Smythe, 'Relating

8,000 individual biographies in individual HTML files that are accessed by means of an internet browser.

For the sake of completeness, one also might mention, in the context of biographical catalogues, a whole category of pseudo-prosopographical databases created by genealogical enthusiasts, many of whom have demonstrated that their families are descended from, for example, Ruricius of Limoges.[31]

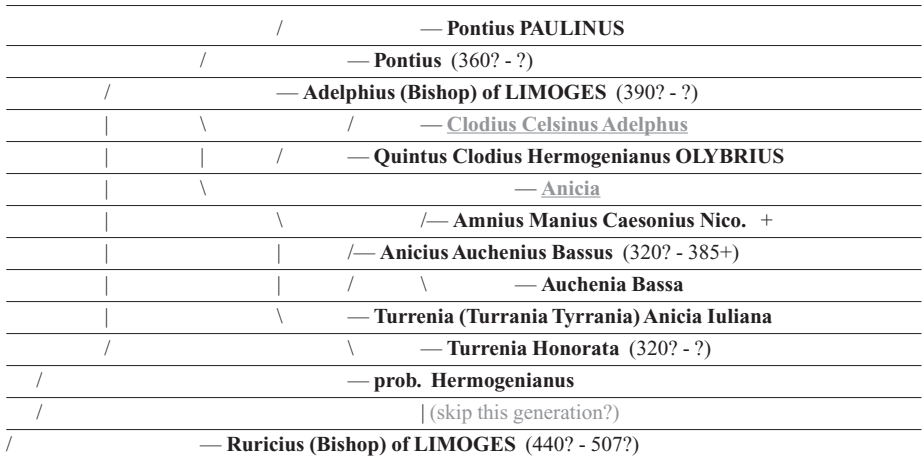
.

SAMPLE ENTRY FROM A TYPICAL GENEALOGICAL DATABASE

The PEDIGREE of Ommace
aka daughter of Ruricius

Poss. HRH William's 40-Great Grandmother. *Poss. Agnes Harris's 33-Great Grandmother.*

Husband **Rusticus (Saint; Bishop) of LYONS**
Children: **Sacerdos (Bishop) of LYONS ; Artemia of LYONS**



In general, biographical catalogues such as these are not only searchable on a just few fields, if not just the 'Name' field, but also have limited or no ability to create subsets of individuals based on shared characteristics. So let us now turn to more complex prosopographical databases and consider the manner in which they implement the methodologies discussed above.

prosopography to the Byzantine Empire: technical details of the computerisation', International Congress of Byzantine Studies at Copenhagen (August 1996): http://www.kcl.ac.uk/ humanities/cch/PBE/TechDetail.htm; Averil Cameron, 'Prosopography of the Byzantine Empire' in *The British Academy Review*, 4 (2000): http://www.britac.ac.uk/pubs/review/04-00b/index.html; Dion C. Smythe, 'Prosopography of the Byzantine Empire', in *DRH99: Selected papers from Digital Resources for the Humanities 1999*, ed. M. Deegan, H. Short (London, 2000), pp. 75-81.

[31] E.g., http://freepages.genealogy.rootsweb.com/~jamesdow/s099/f897937.htm

B. Relational Databases

Since the 1990s, nearly all mainstream PDBs have been constructed using some version of the relational database model,[32] which, as noted above, is based upon tabular databases (also known as 'files', 'datafiles', 'datasets', or 'tables') that contain rows of names (known as 'records', 'entries', 'entities', or 'tuples') each of which has associated with it columns of various categories of data (known as 'fields', 'attributes', or 'domains'). Typical fields include 'Sex', 'Occupation', 'Place-of-Residence,' 'Religion', and so on. More complex data structures include more detailed kinds of categories, such as 'Property-Owned', 'Legal-Activity', 'Education', 'Cause-of-Death', etc.

DATABASE TERMINOLOGY FOR RELATIONAL DATABASES

Database	also known as	File, Datafile, Dataset, Table
Row	also known as	Record, Entry, Entity, Tuple
Column	also known as	Field, Attribute, Domain

Some database software, such as dBase, also permits the storage of variable-length text records in a separate file.

1. Source Oriented Databases

Many early prosopographical databases incorporated primary source material that already was heavily structured, that is, for which the categories of information largely defined themselves. This primarily included various kinds of official records, such as census data, tax records, baptismal records, and so on. These databases are almost always of the 'limited' type, and often return output records that are rather cryptic and require some knowledge of the coding method to be interpreted. A few of them are available on the web. One of them is a descendent of one of the very first medieval PDBs, that based on the Florentine *catasto* [tax survey] of 1427.[33] It contains 10,000 records with 25 fields, of which nine are searchable and seven are coded. In this case, the choice of fields was guided by the nature of the source material, hence fields relating to 'Animals', 'Trade', 'Migration', and 'House-Type.'

[32] See, e.g., James Bradley, 'Relational database design and the reconstruction of the British medical profession: constraints and strategies', in *History and Computing*, 6:2 (1994), 71-84.

[33] See David Herlihy, 'Editing for the computer: the Florentine Catasto of 1427', *American Council of Learned Societies, Newsletter*, 22:2 (1971) 1–7; Idem, 'Computer-assisted analysis of the statistical documents of medieval society', in *Medieval Studies: An Introduction*, ed. J. M. Powell (Syracuse, N.Y., 1976), pp. 184–211; Idem, 'Deaths, marriages, births and the Tuscan economy (ca.1300-1500)', in *Population Patterns in the Past*, ed. F. D. Lee (New York, 1977); M. Demonet, Waldimir Sachs, 'Deux exemples d'application de l'informatique en histoire: la statistique générale de la France, le catasto florentin de 1427-1430', *Informatique et sciences humaines*, 20 (1974), 7-19; J. Paul Bischoff, 'Reformatting the Florentine Catasto Data for Use by Standard Statistical Analysis Programs', *Computers and Medieval Data Processing*, 11.2 (1982), 56; http://www.stg.brown.edu/projects/catasto/overview.html.

INFORMATION RETURNED FROM THE 'CATASTO OF 1427' WEBSITE

Online Catasto of 1427
Edited by David Herlihy, Christiane Klapisch-Zuber, R. Burr Litchfield and Anthony Molho
Catasto SQL Search Interface
Search for records where:
field
field
field
Return the following fields:
To add additional variables to view in the first results screen, highlight them in the variable list above while holding down the [Control] Key (PC) or Command/Apple Key (Mac)
Sort by ; subsort by
Limit the result set to records
Result Set Length, 2; Limit = 60

namepatronymicfamily
namex AGNOLAPIERODELLAGRAMM
LISAROBERTOCANIGIANI

The even more extensive York-Minster Database, containing over 250,000 entries relating to people connected to Yorkshire between 1200 and 1500 and drawn from primary sources including wills, freeman's lists, religious recorder charters and taxation records., returns rather cryptic lists of names and fields such as 'Name', 'Sex', 'Place', 'Date', 'Occupation', and 'Source-Location'.[34]

SAMPLE OUTPUT FROM 'YORK-MINSTER DATABASE'

REPORT ON SEARCHES FOR THE SURNAMES OF RUSHWORTH / RISWORTH AND THEIR VARIANTS:
RISHWORTH,JOHN,M,COLAY,NOJOB,1459,WILL,YASRS/6*
RISHWORTH,JOHN,M,HYPPERON,NOJOB,1475,WILL,YASRS/6*
RISHWORTH,JOHN,M,HALIFAX,NOJOB,1459,WILL,YASRS/6
RISHWORTH,JOHN,M,HALYFAX,NOJOB,1475,WILL,YASRS/6
RUSSHEWORTH,OLIVER,M,PONTEFRACT,NOJOB,1471,WILL,YASRS/6
RISHWORTH,ROBERT,M,PONTEFRACT,NOJOB,l5l9,WILL,YASRS/11
RISHWORTHE,CHRISTOPHER,M,CROSTON,NOJOB,l538,WILL,YASRS/11
RISHWORTHE,WILLIAM,M,FERNELEY TYAS,NOJOB,l534,WILL,YASRS/11

2. Comprehensive Multi-Source PDBS

The greatest future promise of PDBs lies in the construction of more sophisticated and comprehensive databases including a broad range of persons, constructed from a multiplicity of sources, and permitting searching on a multiplicity of fields. Some

[34] http://www.rushworth.com/club/york.html; mailto:alan.tupman@virgin.net.

propose to incorporate pre-determined groups of individuals, defined by some relatively narrow combination of date, location, and role-in-life, whereas others intend to include 'all the people' who lived during more extensive periods of time. These databases are by their very nature open-ended insofar as there is a great wealth of source material to draw upon and there is always the possibility of new source material coming to light. Establishing what the population to be included is to be has a direct influence on how large the end result will be. Most have been constructed using a person-based, single-table relational database structure. More recently, PDBs also have been created using a large-scale multi-file relational structure, and there also have been experiments with a non-person-based data storage model.

Some multi-source databases limit themselves to particular geographical areas of the ancient world. For example, The 'Prosopographical Database of Cultic Personnel in Ancient Rome' proposes to include about 2,000 entries of Roman religious personnel from 300 BC until AD 499.[35] It has searchable fields relating to information such as 'Sex', 'Literary-Activity', 'Offices', and 'Date-of-Birth.' Output entries are created by concatenating the information in the several fields.

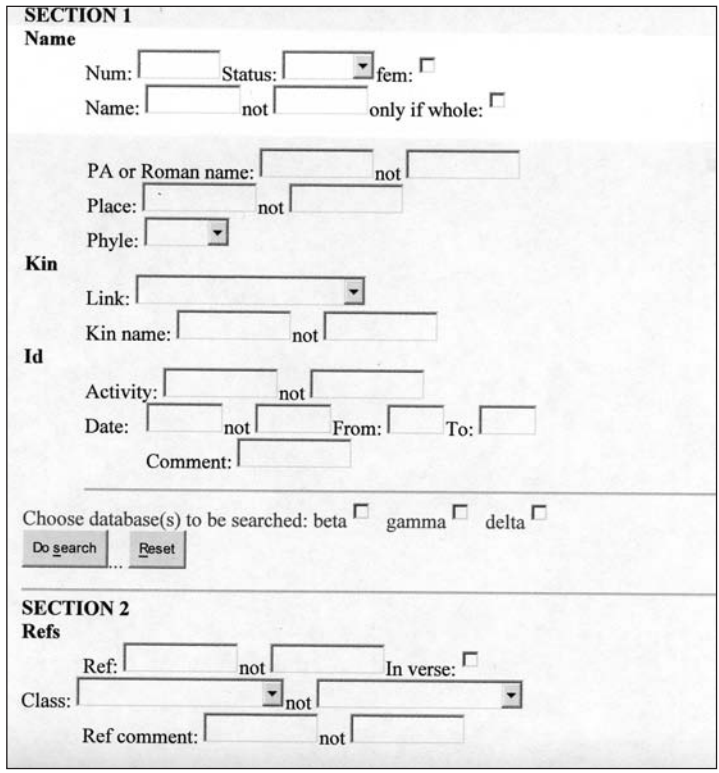
SAMPLE OUTPUT FROM 'DATABASE OF CULTIC PERSONNEL'

Aemilius M. f. M. n. Lepidus, M. (2)
RE Aemilius 68. KP Lepidus I.4. - Bardt P 22, Pon m 13; MRR 1,329. 390; Szemler P 12. M. Pa. Around 230-152 B. C.± - Pontifex since 199, from 180 pontifex maximus. Already 201 legate Lepidus became 193 curule aedile, 187 and 175 consul, 179 censor. Afterwards princeps senatus. Religious activities: Taking down of votive weaponry to finance the renovation of the temple of Iuppiter Capitolinus (as censor 179, Liv. 40,51,3). Preformulates the prayer of the *vota decennalia* 172 (Liv. 42,28,8 f.). 156 the roof of his office-house was thrown into the Tiber by lightning (Obs. 16). Speaking for his college he denied the censor the right to dedicate a statue of Concordia in 152 correctly (Cic. *domo* 136). RRC 419/2 (pontifex maximus); Liv. 32,7,15 (cooptation); 40,42,12 (election as pontifex maximus); *per.* 48 (death). Münzer 1920:170-8. 201 f. 410; Draper 1988:246. 1 For the year of birth s. Münzer 1920:171. 177; for the year of death additionally Bardt.

The 'ATHENIANS', database, available at 'Website Attica', is a partial computerization of about 10,000 of the 100,000 persons (with names beginning with the letters beta, gamma, and delta) taken from the 'Persons of Ancient Athens.'' It is based primarily on epigraphic sources. Using pull-down menus based on fields relating to 'name', 'status', 'place', 'phyle', etc., it returns coded lists of names of persons meeting input criteria relating to data such as 'Sex', 'Location', and various forms of names.

[35] http://www.uni-potsdam.de/u/klassphilol/priedv.htm.

INPUT SCREEN FOR 'WEBSITE ATTICA'



SAMPLE OUTPUT FROM 'ATHENIANS' DATABASE

322260 DHMULOS (AURHLIOS) (PAND)
in catalogue of areopagitai & others, 180-192?p. Status A+R.
1.1 II 2339, line 20 (cat arch).
Pan[dioni'dos] / 'Areopagei=tai // [.5.] 'Hdu'los
< 1.2 AEph 1950 p31 16, line 22
Pan[dioni'do]s ~ Au'r Dhmu'los
= 1.3 H 27 p44, line 22
< 1.4 H S12 p166, line 93
[Pandioni'dos] / 'Areo[pa]gei=tai ~

The 'Biographical Database of Late Antiquity' (BDLA), on the other hand, plans to incorporate all the persons attested as living in the Mediterranean and western Asian worlds between AD 250 and 750. It currently contains over 27,000 individuals and is searchable on over 70 fields.[36]

[36] R. Mathisen, 'Creating and using a biographical database for Late Antiquity', *History Microcomputer Review*, 5:2 (1989), 7-22.

DATA FIELDS FOR BIOGRAPHICAL DATABASE FOR LATE ANTIQUITY

	Field Name	Data Type	Field Length	Field Type		Fiel Name	Data Type	Field Length	Field Type
01	NAME	Character	40	S	34	MARITAL	Character	1	A
02	NAMESTAND	Character	15	S	35	CHILDREN	Character	1	A
03	GENS	Character	15	H	36	RELATIVES	Character	1	A
04	ALTSPELL	Character	1	H	37	RELIGION	Character	1	A
05	DUPLICATE	Character	1	H	38	CLASSBIRTH	Character	1	S
06	NATIONALTY	Character	1	S	39	CLASSDEATH	Character	1	A
07	AREAORIG	Character	1	S	40	SAINTHOOD	Character	1	A
08	DIOCESORIG	Character	1	S	41	SOCIALACT	Character	1	A
09	PROVORIG	Character	20	S	42	PROFACT	Character	1	A
10	PLACEORIG	Character	20	S	43	LEGAL	Character	1	A
11	PLACECALC	Character	20	C	44	ECONSTATUS	Character	1	A
12	DATEACTIVE	Character	10	S	45	ECONTRANS	Character	1	A
13	BIRTHDATE	Character	5	S	46	PROPERTY	Character	3	A
14	DEATHDATE	Character	5	S	47	EDUCATION	Character	1	A
15	DEATHDAY	Character	4	S	48	WRITING	Character	1	A
16	DEATHAGE	Numeric	3	S	49	LETTER	Character	1	A
17	DEATHCAUSE	Character	1	S	50	TRAVEL	Character	1	A
18	HEALTH	Character	1	A	51	RELIABILTY	Character	1	A
19	OCCUPATION	Character	2	A	52	EPITAPH	Character	1	A
20	ACTIVITY	Character	65	A	53	ANCSOURCE	Character	1	H
21	POSITION	Character	30	A	54	ANCIENTREF	Character	30	H
22	PLACE	Character	20	A	55	MODERNREF	Character	30	H
23	DATE	Character	10	A	56	ADDITREFS	Character	1	H
24	POSITION2	Character	30	A	57	COMMENTS	Character	65	H
25	PLACE2	Character	20	A	58	DBSOURCE	Character	1	H
26	DATE2	Character	10	A	59	BIOEXIST	Character	1	H
27	POSITION3	Character	30	A	60	BIOGRAPHY	Memo	10	A
28	PLACE3	Character	20	A	61	DATENENTRY	Character	6	H
29	DATE3	Character	10	A	62	DATEUPDATE	Character	6	H
30	ADDITOFFIC	Character	1	A	63	DATEBEG	Numeric	3	C
31	MILSERVICE	Character	1	A	64	DATEEND	Numeric	3	C
32	HONORS	Character	1	A	65	DATESORT	Numeric	3	C
33	SEX	Character	1	S					

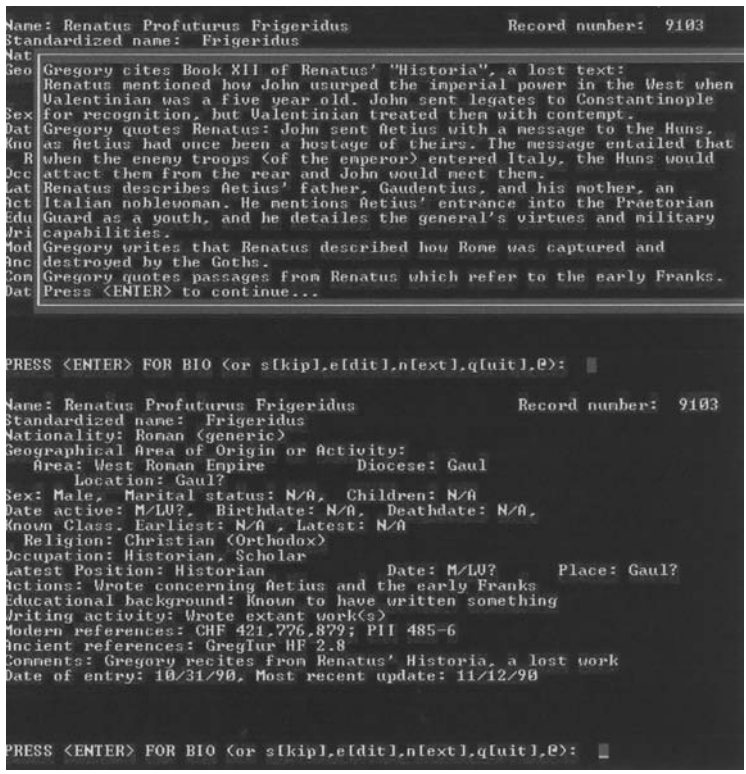
SEARCH CODES FOR OCCUPATION FIELDS OF 'BDLA'

OCCUPATION: The individual's occupation, activity, or role in life.

X 'Independently Wealthy' (aristocrats without offices)

C Civil Service incl. Rulers
M Military
U Civil AND military service
E Church cleric
B Church AND state office
L Literary figure
G Grammar, Rhetoric &c Teacher
H Historian, prof. scholar
S Student
P Philosopher
W Wife
C Child
Q Entertainer (mime, actor, gladiator, charioteer, etc.)
O Owner of large estates
R Merchant, Shopkeeper
A Artisan, Craftsman
Y Advocate, Lawyer
D Medical Doctor, physician
K Workman, laborer
F Peasant Farmer
N Servant
V Slave
T Tenant farmer, colonus
J Jail Inmate
Z Beggar

SAMPLE OUTPUT FROM 'BDLA', WITH 'BIOGRAPHY' PAGE



a. Multi-File Relational Model

The preceding projects all are based, fundamentally, upon storing the information for each person all in a single record in a single database, accompanied, perhaps, by an ancillary file for the storage of overflow data. Other projects make use of subsidiary databases to store certain categories of information. The 'Prosopographica Ptolemaica' project, for example, which includes over 100,000 records, stores information on 'Texts', 'References', 'Persons', 'Families', 'Offices', and 'Name-Variants' in different files, all indexed on person-name.[37]

DATAFILE CONTENTS FOR 'PROSOPOGRAPHICA PTOLEMAICA' PROJECT

1. TextFile: 9,001 entries
2. ReferenceFile: 32,588 entries
3. PersonFile: 21,321 entries
4. FamilyFile: 13,720 entries
5. NameFile: 7,147 entries
6. FunctionFile: 670 entries.

[37] http://prosptol.arts.kuleuven.ac.be/

It also is available on a website, where menu-drive searches can be input, and records on individual persons are returned.

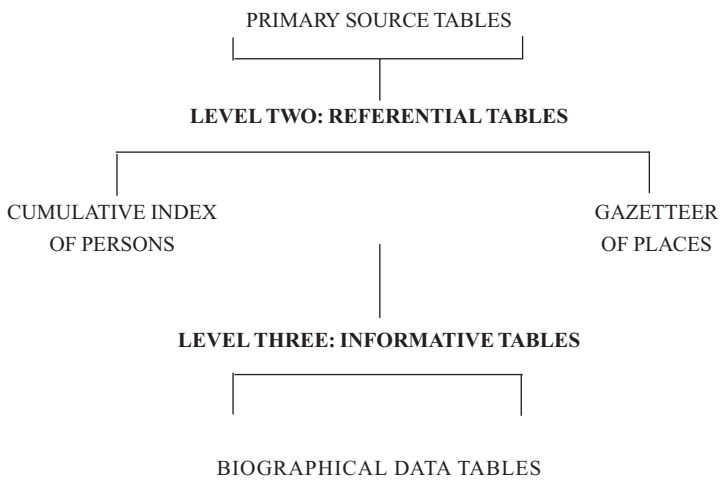
SAMPLE OUTPUT FROM 'PROSOPOGRAPHIA PTOLEMAICA' DATABASE

Record details:	
pstnam	**Ptolemai'oß**
pstnamlat	**Ptolemaios ho Keraunos**
psex	**1.Man**
peth	
presid	
pfirst	**BC**
plast	**BC**
pperiod	**Ptol**
ppp	**14539***
pbib	
functions	**royal family**

Find corresponding ref-records

Likewise, the 'Continental Origins of English Landowners' (COEL) database, stores information on sources, names, places, and so on in 3 levels of tables with several searchable fields.[38]

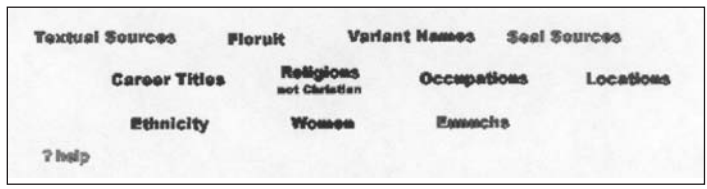
STRUCTURE OF COEL DATABASE
LEVEL ONE: PRIMARY SOURCE TABLES



[38] Keats-Rohan and Thornton, 'COEL and the Computer', *Medieval Prosopography*, 17:1 (1996), p. 256.

The published version of *PBE I* includes, along with the 8,000-plus individual biography entries, a set of pre-formatted indexes, based on catetories of information ranging from 'Floruit' to 'Ethnicity' to 'Eunuchs,' that permits sub-sets of persons to be selected out who meet combinations of search criteria.

***PBE I* SEARCH CATEGORY SELECTION SCREEN**



For example, a search of 'Career Titles' brings up an index that can then be clicked upon to produce a list of persons holding that title.

***PBE I*: INDEX OF CAREER TITLES AND RETURNED PERSONS**

P, Palatinos, Papal, Papas, Papias, Parakoimomene, Parakoimomenos, Paraphylax, Paroikonomos, Patriarch, Patriarchikos, Patricius, Patrikia, Patrikios, Pech

Palatinos

Seon 1 (MIX)
Theodoros 153 (VI/VII)
Theodoros 191 (VII/VIII)
Theodoros 348 (VII)

b. Decentralized Biography Model

In recent years, an alternative method for structuring PDBs has been developed, in particular, by the Centre for Computing in the Humanities at King's College, London in the 'Prosopography of the Byzantine World' (PBW)'' (formerly the 'Prosopography of the Byzantine Empire' [PBE]) project (the first to essay this methodology);[39] the 'Prosopography of Anglo-Saxon England' project;[40] and the 'Clergy of the Church of England' (CCE) project.[41] In this model, individual records with dedicated fields are not created for each individual. Instead, each person is assigned a unique identification key (e.g. a name and a number). This key is then associated with information-bites (called 'factoids' by their creator Dion Smythe) in any number of other databases, e.g. a 'Locations' database, a 'Titles' database, an 'Occupations' database, an 'Ethnicity' database, a 'Variant-Names' database, and so on. As noted on the project website: 'The

[39] See John Bradley, Harold Short, 'Using formal structures to create complex relationships: The Prosopography of the Byzantine Empire – a case study', in. *Resourcing Sources: The Use of Computers in Developing Prosopographical Methodology,* ed. K. S. B. Keats-Rohan, Prosopographica et Geneologica vol. 7 (Oxford, 2002): http://pigeon.cch.kcl.ac.uk/docs/papers/pbe-leeds/index. html; and John Bradley, Harold Short, 'Texts into databases: The Evolving Field of New-style Prosopography' (2003): http://pigeon.cch.kcl.ac.uk/docs/papers/georgia1/.

[40] http://eagle.cch.kcl.ac.uk/pase/index.jsp; http://www.kcl.ac.uk/humanities/cch/pase.

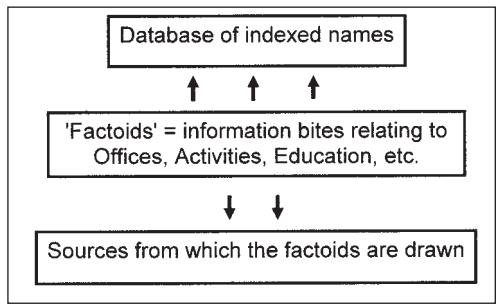
[41] http://www.kcl.ac.uk/humanities/cch/cce.

set of factoids represents a systematic and structured view of what are regarded as key types of personal information [including]:

Authorship (Person A wrote a particular document)
Education (Person A was educated by person B)
Event (Persons A, B, C took part in a particular event , e.g. a battle)
Occupation (Person A was engaged in a particular occupation)
Office (Person A held a particular office)
Personal relationship (Person A and person B had a particular relationship between them)
Personal information (e.g. Person A had red hair, was a saintly man, …)
Possession (Person A was owner of a particular object)
Recorded name (Person A was named in a particular way in this source)
Status (Person A held a particular status in their society)
Transaction (Persons A, B, C were involved in the exchange of ownership of something).

In addition, the full-text sources of the factoids can be accessed in additional databases containing texts of the primary sources.

RELATION OF 'FACTOIDS' TO NAMES AND SOURCES



Thus, each person can have associated with them any number of locations, offices, activities, and so on. In order to create a 'biography' for any particular individual, all of the factoids relating to that person are concatenated.

REASSEMBLY OF FIELDS FOR 'BASILIOS 7' FROM PBE I DATABASE[42]

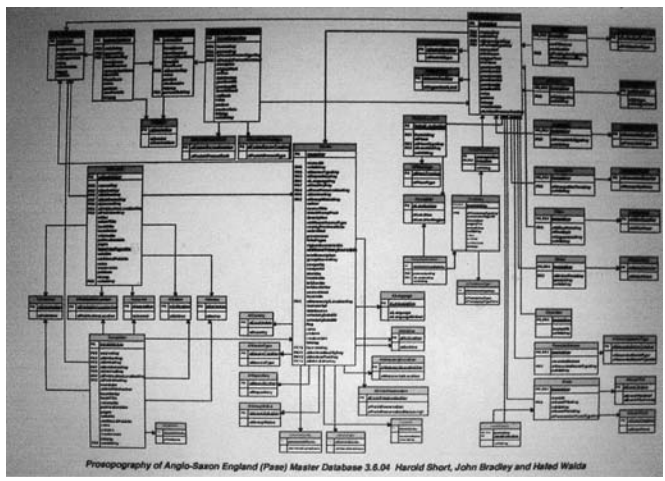
Basilios 7	*Prosopography of the Byzantine Empire*
Sex	M
Floruit	E/L IX
Dates	813 (n.) / 886 (ob.)
PmbZ No.	832
Variant Names	Basileios

[42] http://www.kcl.ac.uk/cch/PBE.

Ethnicity	Macedonian; Armenian
Locations	Kepoi (Thrakesioi); Macedonia (property); Peloponnesos; Hagios Diomedes (Monastery of, Constantinople) (topographical); Constantinople (officeplace); Adrianoupolis (Macedonia) (residence); Bulgaria (residence); Constantinople (residence); Adrianoupolis (Macedonia); Bulgaria; Constantinople; Adrianoupolis (Macedonia) (birthplace)
Occupation	Farmer
Titles	Patrikios (dignity); Augustus (office); Basilikos protostrator (office); Basilikos strator (office); Emperor (office); Parakoimomenos (office); Protostrator (office)

The decentralized biography approach, as manifested, e.g., by the PASE database, can result in a forbidding array of tables and files, many of which can access each other.

DATA STRUCTURE FOR 'PASE' DATABASE



The PASE database has been on-line since May 2005, and can be searched based on several different categories of information and on many different values for each category. By navigating through the menus, and constructing detailed search criteria, one can select out groups of people meeting stated criteria, the primary goal of any good PDB.

SEARCHABLE FIELDS IN 'PASE' DATABASE

Select Index ››
Persons
Sources
Status
Locations
Events
Offices
Occupations
Relationships
Possessions

EXAMPLES OF VALUES FOR 'STATUS' IN 'PASE' DATABASE

Index of Statuses
Accuser
Adolescens
Apostate
Apostle
Bretwalda
Burgess
Burhwaru
Captive
Ceorl
Child
Ciuis
Cliens
Clito
Cnapa
Comes
Comitatus
Comitissa
Commoner
[etc.]

EXAMPLE OF OUTPUT FROM 'PASE' DATABASE

Cancellarius
Eanfric 1 - l viii (*Witness to spurious charter of St Denis, '795'*)
Wolsin 1 - e/m x (*Cancellarius in spurious charter, '924x939'*)

The multi-file relational model, combined with the decentralized biography approach, is not, however, without potential caveats that ought to be taken into account by anyone essaying to implement this methodology. For example,

1. The complexity of the data structure means that it is difficult, if not impossible, to implement by anyone without expert computer skills or an expert staff;
2. The multiplicity of sub-databases and lack of core biographies make it difficult to export material to other software or to integrate it with other PDBs without using special programming;

3. The lack of 'base-level' unity provided by individual 'person' entries can increase the likelihood that significant information for an individual could be omitted when the information-records are combined. For example, in *PBE I*, the recombined data for Byzantine emperor Basil VII lists for 'Occupation' only 'Farmer,' something that would never have happened with an individual lemma[43] (Martindale's associated stand-alone lemma immediately describes him as 'Emperor').

4. The lack of a 'home entry' to refer to also can make it difficult to determine whether occurrences of the same name represent the same person or different persons. One only has 'virtual persons' created by recombining a multitude of individual information records. But how is a staff member – perhaps a transient student – to decide which person a new piece of information is to be associated with without a core biography for each person to refer to? One can end up in a process of infinite recursion if the persons with whom a multitude of information-records can be associated are always up for grabs.

5. In general, one might worry that we have come full circle, with some projects going back to databases that have the appearance of being dependent on a complex software configuration and a particular staff who understands how it works.

VI. Where are All the PDBs?

This general survey covers only some of the major PDB projects undertaken during the past thirty years. In point of fact, not many have been left out. For the classical and medieval worlds, we have no more than ten grand-scale (that is, with over 2,000 person entries) functional PDBs to show for the past 30 years. Why is this the case? The preceding examples demonstrate that in order to create a large scale, complex, fully functional PDB it is necessary that several desiderata be met:

1. Lots of funding – my own experience with the Biographical Database for Late Antiquity project suggests that each entry costs about $12.00 – and that was 10 years ago.
2. An effective Principal Investigator. It is essential that a PI not only be knowledgeable about both computer issues and prosopographical issues, but also take a hands-on approach to every step of the process.
3. A dedicated, long-term, knowledgeable staff, which can be very difficult to find, given that students, the most ready source of Project Assistants, keep cycling in and out of a program.
4. A rigorous implementation of database standards and procedures: there cannot be any shortcuts; and
5. Lots of time.

All of which is beginning to sound like building the pyramids, and helps to answer the question of why we do not have more PDBs on the grand scale. Creating an effective PDB is not for the faint of heart. It requires at least a modicum of technical skill; it requires a lot of work 'up front' before data can be entered; it almost always requires funding; and, not least, it requires perseverance.

[43] [Lemma, pl. lemmata: a headword in a dictionary or similar list.]

But there also might be another answer to the question of the missing PDBs. They might be hiding. An unscholarly survey undertaken at the Prosopography Conference for which this overview was written suggests that nearly all prosopographers have created their own personal databases for their own research, some using database software, but many using EXCEL or just WORD. My own students make them as a matter of course. I would suggest that we also are living in an age of small, personal databases created to satisfy specific research agendas and which rarely get counted in 'surveys of PDBs'. Several of these may lurk behind some of the prosopographical projects cited in Katharine Keats-Rohan's 'Directory of Prosopographical Research'.[44]

VII. Concluding Recommendations

This survey might be concluded with some suggestions for those who might be considering creating their own database.

1. 'Keep it simple'

For most projects, especially personal PDBs, a database structure based on a single tabular file with individual person entries remains the most practical:

1. They require little or no programming ability, and do not need time-consuming construction of specialized data storage formats;

2. One can begin entering data essentially immediately;

3. Any disadvantages that accrue as a consequence of lack of space in the data structure for pieces of information for very well attested persons can be dealt with by (1) using multiple fields for the same data category; (2) including that information in a variable-length BIOGRAPHY field, (3) including it in a very limited number of sub-databases, e.g. an occupations or family database, or (4) simply duplicating an entry for these persons to accommodate the extra information. Lack of enough space for exceptionally well documented persons, therefore, need not be an argument against using a tabular format;

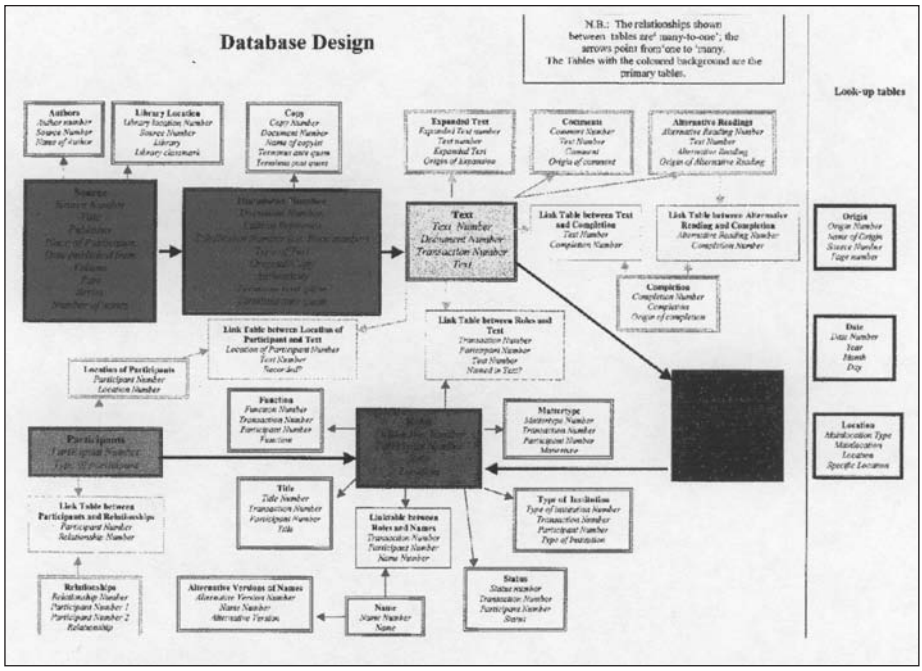
4. A database in a single-table format is always exchangeable with other databases, software, and hardware. It can be exported to another computer or different type of software with a minimum of trouble. For this purpose, only the relational model really works without going through some kind of cumbersome, and possibly erroneous, conversion process. Flat-file relational databases will never be obsolete. For example, my 1974 computer card file is easily converted directly into an EXCEL file:

FL EUGENIUS	1	1	G	C		I	G	K	392	N		3	A	A	392	0	2	3							VC 385	G	
EUETERIUS	1	2		E					LIV			6														L	L
EUDOXIUS	1	2		G					382	0	1	4														E	
EUDOXIUS	1	2		G					379	0	4	C														E	
EUCHROTIA	1	B			B		X	I	385	0	0	B													MAXIMUS	L	
FL EUCHERIUS	1	U		A		I	A	H	377	0	2	U	H	B	381	0	0	U	X	H	392	0	0	2		G	
EUCHERIUS	1	U		A			A	R	397	0	8	1				E										L	
EUCHARIUS	1	U		A		S	A	X	383	0	0	J														C	

[44] Accessible from http:// prospography.modhist.ox.ac.uk/directory.htm

A cross-indexed multi-file database might sound nice, but it is not something one should try at home unless one has lots of funding, lots of time to bone up on database theory, and a dedicated staff.

STRUCTURE FOR MANCHESTER 'PATRONAGE' DATABASE[45]: DON'T TRY THIS AT HOME!



2. 'Don't recreate the wheel'

After 30 years, there is a tremendous amount of literature about what works and what does not. Anyone proposing to construct a large PDB would be well advised not to try to re-invent the wheel.

3. 'Pay me now or pay me later'

There is no substitute for advance planning. Revising a database structure, and its contents, is much more time consuming than revising a manuscript. Get that data structure and coding format down pat, and test it thoroughly, before beginning to enter data. If you leave something out, it will come back to haunt you later. In general, PDBs always expand with respect to the number of fields, the number of codes, the number of sources included, and the number of persons. What we thought was unimportant, or what we did not consider at all, will always come back to haunt us.

[45] http://www.art.man.ac.uk/cla/patronage.htm.

4. 'No data left behind'

In particular, most modern PDBs allow searching on only a relatively limited set of criteria. Increasing the number of categories of information stored would greatly increase the number of questions that could be asked. Doing so, of course, violates one of the general cautions that some have set down for database construction: not to have too many fields. But as demands upon PBDs grow, and users move beyond the standard queries involving gender, social class, and offices held, more subtle questions will arise. So, if a category of information even seems like it might be useful, put it in. You can always take it out later

5. 'Don't jump the gun'

We must unlearn some of the things we have learned based on other scholarly experience. In particular, multi-tasking does not work. The definition of data structure and coding format must be as complete as possible before serious data entry begins.

6. 'How much longer?'

If done properly, everything will take about five times as long as anticipated – or more!

7. 'You just can't get good help'

In the case of large-scale databases that require staff assistance, think very carefully about who will be doing the data entry.

8. 'Getting the news out'

Far and away the best way to make a database accessible to users is to make it available on the World Wide Web.

9. Backup, backup, backup

In sum, creating a PDB is not for the faint of heart. It is full of frustrations, backtracking, and missed deadlines. PDBs need to be nurtured, coaxed, and given lots of second chances. PDBs are a labour of love. So perhaps the best advice to give to those in the throes of creating a PDB is simply to 'hang in there'.

Suggested Reading

Althoff, G., 'Zum Einsatz der elektronischen Datenverarbeitung in der historischen Personenforschung', *Freiburger Universitätsblätter*, 52 (1976), 17 32

'Ancient Studies, New Technologies: The World Wide Web and Scholarship in Ancient, Byzantine, and Medieval Studies' (2000-04), http://www.cisat.jmu. edu/asnt3/.

Barman, J., R. Barman, W.T. Kirshaw, W.T., 'Prosopography by computer: the development of a data base', *Historical Methods Newsletter*, 10 (1977), 102–8

Bradley, John, Harold Short, 'Using formal structures to create complex relationships: The Prosopography of the Byzantine Empire – a case study', in. K.S.B. Keats-

Rohan, ed., *Resourcing Sources: The Use of Computers in Developing Prosopographical Methodology,* Prosopographica et Geneologica vol.7 (Oxford, 2002): http://pigeon.cch.kcl.ac.uk/docs/papers/pbe-leeds/index. htm

Bradley, John, Harold Short, 'Texts into databases: the evolving field of new-style prosopography' (2003): http://pigeon.cch.kcl.ac.uk/docs/papers/georgia1/

Bulst, Neithard, 'Prosopography and the computer: problems and possibilities', in P. Denley, S. Folgelvik, C.Harvey, eds., *History and Computing II* (Manchester, 1989), pp.12–18

Averil Cameron, 'Prosopography of the Byzantine Empire', in *The British Academy Review*, 4 (2000): http://www.britac.ac.uk/pubs/review/04-00b/index.html

Dawson, John, 'ACAD – A Cambridge alumni database', *Computers in Genealogy*, 7.8 (2001), 363–75

Glenisson, J., 'Prosopographie et informatique', in L. Fossier, ed., *Informatique et histoire médiévale* (Rome, 1977), pp.227–9;

Goudriaan, Koen, Kees Mandemakers, J. Reitsma and P. Stabel, eds., *Prosopography and Computer: Contributions of Mediaevalists and Modernists on the Use of Computer in Historical Research* (Leuven, 1995).

Greenstein, D.I., *A Historian's Guide to Computing* (Oxford, 1994)

Harvey, Charles, Jon Press, *Databases in Historical Research. Theory, Methods and Applications* (New York, 1996);

Keats-Rohan, K. S. B., David Thornton, 'COEL and the computer: towards a prosopographical key to Anglo-Norman documents', in R.W. Mathisen, ed., 'Late Antiquity and Byzantium,' special issue of *Medieval Prosopography*, 17:1 (1996), 223-62

Martindale, J.R., ed., *The Prosopography of the Byzantine Empire I: (641–867)* (Aldershot, 2001).

Mathisen, Ralph W., 'Medieval prosopography and computers: theoretical and methodological considerations', *Medieval Prosopography*, 9 (1988), pp. 73–128;

Neiske, Franz, 'Die Erforschung von Personen und Personengruppen des Mittelalters mit Hilfe der elektronischen Datenverarbeiten', in K.F. Werner, ed., *Medieval History and Computers* (Munich, 1981), pp. 77–110;

Smythe, Dion C., 'Prosopography of the Byzantine Empire', in M. Deegan, H. Short, eds., *DRH99: Selected papers from Digital Resources for the Humanities 1999* (London, 2000), pp. 75-81

Thaller, M., 'Methods and techniques of historical computation', in P. Denley, D. Hopkin, eds., *History and Computing* (Manchester, 1987), pp.147-56

Townsend, Sean, Cressida Chappell, Oscar Struijvél, *AHDS Guides to Good Practice: Digitising History*. (Oxford: Oxbow, 1999): http://hds.essex.ac.uk/g2gp/digitising_ history/.