SURNAME GROUPING

PURPOSE OF GROUPING

Similar names with different spellings will invariably appear in any list of surnames. Variations in spelling are important in modern times, where almost everyone is able to read and write and people will strenuously defend their own particular spelling of their surname. Some people go out of their way to spell their name differently. However in the times when few of the populace were literate, the particular way in which a name was spelt was much more to chance and the whim of the scribe who wrote the name and was very often of little consequence to the person themselves. Recognising these variations of spelling has been a problem for scholars and historians for many years. However most of the vagaries of the spelling are quite clear and obvious to literate and educated researchers. Indeed this is an area that the human brain copes with especially well, not so the computer which sees every separate spelling as a separate surname, unless it is told different. A computer database therefore needs assistance, from the human experience, to devise a method by which the differing formations of a particular surname can be associated together.

Grouping of surnames is carried out therefore to enable a computer to search a database and then to bring together the variations of surnames that may have a common base. Surnames that are so grouped together and brought together in a search, should not be taken as necessarily belonging to the same family or individual. Grouping is carried out merely to bring to the notice of the researcher all those surnames that may belong together.

It is especially important that the spelling of the surname in the database is preserved in its original form. The technique used to support surname grouping comprises; allocating a code to each surname variation and then carrying out the database search on this code and not on the original surname. The researcher is able to use the original surname version to make a considered judgement which names to utilise and which to disregard.

In the database of 17C Thame a special surname code, called a FISC, is created that will accommodate the variations of spelling and diversity of pronunciation that were prevalent in the 17C. Before explaining how the code is created, it is probably best to understand how the surnames initially evolved into separated spellings. Considering the names stored in a computer database there are three main areas where the differences may have occurred;

COMPOSITION, TRANSCRIPTION AND KEYING-IN.

Composition.

Scribes would be employed to carry out most of the written work and the dialect and poor pronunciation of the subject person may cause a name to be mis-heard, and thus *Ayres* may have been heard and written down as *Eares*. Scribes differed in their learning and thus also in the way that they spelt many names; Crewes is the same as *Crues*, *Bue* the same as *Bew*. Less able scribes may have written words more phonetically thus *Eustace* may be written as *Ustis*. Even the best scribes would make mistakes and letters may be transposed, *Slater* being written as *Salter*, letters would also be added or left out particularly the letter *"e"*. Consonants may be doubled or singled apparently at the whim of the scribe, *Barnett* and *Cotonn* being typical examples. Although not strictly mistakes, some letters are used as alternatives for other letters and on this basis the letters *"d"* and *"t"* would indicate that *Adkins and Atkins* are the same name. The letters *"ie"* and *"y"* as used in *Clisby* and *Clysbie* and the *"v"* and *"b"* when used in *Nevell* and *Nebell* are further examples of inter-changeable letters. These all add to the problem.

Transcription.

The transcription of names from the old secretary hand is not made easy by the poor quality of some of the handwriting, particularly where the fluidity of the scribe's hand can make vowels a special transcription problem. Vowels can be difficult to differentiate from other vowels, especially the *"a"* and *"e"* hence *Berry* is the same as Barry. The letters "o", "u" and "e" can readily be confused which could mean that *Alnot, Alnut* and *Alnet* are the same name. Groups of small vertical pen strokes known as "minims" are a constant transcription problem, making differentiation between "m", "n" and "u" and even "i" very difficult, if not impossible, thus *Haunces* may well be the same as *Hamce*. Some other letters are liable to be confused in transcription and thus lead to inaccuracy. The letter *"r"* takes the form of a modern letter "c" and so the name transcribed as *Gafecie* is actually *Geferie* and thus is *Jeffery*. The long *"s"* is often transcribed as an *"f"* although in the case of the name *Woodrase* the *"s"* is actually a *"f"* and the name probably is *Woodrofe*. Flamboyantly written capital letters cause their own particular problem,

being particularly difficult to read and transcribe and so it is not surprising that *Randolph* may become *Bardolph* and *Jemott* may become *Irmett.*

Keying-in.

Despite the best efforts of the typist, errors may still occur when data is being keyed in. These errors are difficult to identify, especially as they are very similar in content to other problem areas. Transposition of letters may be attributed to keying-in and thus, *Calcke* is the same as *Clacke*, which is the same as *Clarke, Brain* may be *Brian,* and *Salter* is possibly *Slater.* It is obvious also that these last examples may well have been an inaccurate transcription or a composition problem, in the first instance, or they may have not been an error at all. A keying-in error will occur where the incorrect keyboard letter is pressed and these are most easily identified where an adjacent keyboard letter occurs in a word; Spemser for *Spencer* and *Weaton* for *Weston.* Once a problem is identified as a keying-in error, it can be readily resolved by reference back to the original document or transcription. Unlike many of the Composition or Transcription errors, keying-in errors can be eliminated.

After all these consideration it must always be heeded that the way in which a surname is written may well be the way that the person wished it to be written and thus it is, after all, the correct and distinct surname.

FOUR LETTER INITIAL SURNAME CODE (FISC)

The surname grouping used in the database of 17C Thame is produced by manually creating a four-digit code known as a FISC for each and every surname variant. The surname and its FISC are stored in a table of their own and prior to a database search, an update query is run that populates a *code field* with the FISC for each of the names in the relevant database table.

Taking the first letter of the surname plus the next two consonants and then the last consonant creates the FISC. Vowels (including the letter Y) are ignored, except when they occur as the first letter of the surname. Using this method the name *Painting* has a FISC code of *PNTG* and the names *Bairstone* and *Barstonn* both have a FISC of *BRSN..* The doubling of letters is coped with by always treating double letters as a single letter. Thus*, Allertt* has a code of ALRT.

A search of the database for the name Jemmot is carried out on the FISC code JMT, which will return the names Jemmot, Jemmott, Jemet, Jamot etc.

Distinct Surname Groups.

Where a number of surnames variants can be deduced to be of the same surname base, then these names are considered to form a distinct surname group. The same FISC is allotted to every surname that is deduced to belong to the group, regardless of the letter formation of the surname. The group FISC that is used is normally the FISC that is judged most appropriate or convenient.
Thus: *Eyres* is coded as ARS (Distinct Surname Group *Ayres)* and *Morrice* is coded as MRS (Distinct Surname Group *Morris).* Therefore for the name that has been transcribed as *Irmet* to be returned with a search on the name *Jemmot* (JMT), it will need to be coded JMT.

Special Case Letters.

Certain letter combinations need special consideration when being coded, for instance the names *Wright* and *Write* or even *Rite* need to be grouped together. This type of letter combination is best coded phonetically and so "Wr" is coded as R and "ght" as T, therefore the code for *Wright* is RT and so is *Write* and *Rit*e. However the letter combination "gh" without the letter "t" is generally coded as just as H as in the name *Bough* which codes BH. The case of silent letters must also be dealt with phonetically and hence silent letters are left silent and omitted therefore *Knapp* is coded NP and *Lamb* is coded LM. A full list of special case letters is contained in a table at the end of the text.

Similar Surname Groups.

Some surname groups give particular problems, particularly where variations of the name may have evolved from more than one surname. In these cases it is preferable to include all variants in a single group rather than try to separate them. For example, variations of the name *Bowyer* are impossible to accurately separate from the surname

variants of *Boyer*. Variants could equally well be adjudged to belong to either base name and thus a single group containing all variants, although potentially large, will return all possibilities and therefore allow a database searcher to disregard the undesirable names.

Surname Groups from First Names.

Surname variants that derive from a common first name are all included in one group. The group contains all derivations of that name including suffixes (son) and prefixes (Ap) and so *Davies* and *Davidson* are placed in one group (DVDN) and similarly *Ap Powell* is included in the group *Powell* (PWL).

CONCLUSION

The Thame database has nearly 27,000 surname entries that can be searched in a matter of seconds to find all occurrences of a particular surname group.

The list of surnames in the 17C Thame database currently contains some 3492 variations each with the FISC for that name. The number of different FISCs (i.e. surname groups) derived from these names is 1209. The list of surnames is being continually increased. Whenever a new surname or list of surnames is added to any of the database tables, the names are checked against the list of FISC surnames and any that are not already contained in the list added and a FISC created for the surname.

RULES FOR CREATING A FISC

• The first letter of the surname is used followed by the next two "usable" consonants from the surname and the last "usable" consonant.	Brittain	BRTN
• A vowel is used only when it occurs as the first letter of the surname.	Abats	ABTS
• Double letters are always treated as a single letter.	Allertt	ALRT
• When less than four letters are used to form the FISC unused spaces are left blank	Bennett	BNT

Special Case Letters and Letter Combinations

•	CH	Coded as "C".	Church	CRC
•	CK	Coded as "K".	Cocker	CKR
•	DG	Coded as "G".	Badger	BGR
•	GH	Coded as "G" when it is the first surname letter, afterwards as "H'	Ghost Borough	GST BRH
•	GHT	Coded as "T"	Brighton	BRTN
•	KN	Coded as "N".	Knappe	NP
	LK	Coded as "K"	Walker	WKR
•	Mac/Mc	Coded as "M".	McDonald	MDND
•	MB	Coded as "M".	Wycombe	WCM
•	MP	Coded as "M".	Bampton	BMTN
•	PH	Coded as "F'	Phillips	FLPS
•	RH	Coded as "R".	Rhodes	RDS
•	SH	Coded as "S".	Ashbourne	ASBN
•	SPR, STR, SCR	Coded as "SP" "ST" "SC" (The "R" is not used)	Springall Stringer	SPNL STNR
•	TCH	Coded as "C'	Mitchell	MCL
•	TH	Coded as "T'	Thomas	TMS
•	WH	Coded as "W"	White	WT
•	X	Coded as "KS" or just "S" when it is the last letter	Dixon Wilcoxe	DKSN WLCS
•	Y	Treated as a vowel except when it is the first or last letter of the surname.	Yeates Monday	YTS MNDY
•	Z	Coded as "S".	Zester	SSTR
	oe	At the end of a word coded as "W"	Parsloe	PRSW
	ie	At the end of a word coded as "Y"	Pettie	PTY

Thame History Research Group
1 September 1999