Big family problems? Spread them out

By Allin Kingsbury

Many family history programs are now into their second and third generations, and seem to do everything except find the data. But there is one problem not easily solved by PAF or any other family history program. I used the Excel spreadsheet program to attempt to solve the problem. The spreadsheet proved effective, and so I shall pass my experience along for any who may benefit.

The spreadsheet was designed as an accountant's tool and has become an extremely popular personal computer program. The flexibility and versatility of the spreadsheet program make it useful for many nonfinancial applications. There are two features that make it especially helpful for manipulating genealogical data:

Related data can be linked in rows and columns.
Data can be sorted by date or by any other attribute.

1. Linking related data

My problem was that I had a large number of individuals with similar names and birthdates. This made finding the right individual in the PAF program slow and difficult. By sorting the data in the spreadsheet, I could compare data of similar individuals on the same computer screen and select the dates and places that were possible matches for the individual I was looking for.

I first used the spreadsheet to compile the large number of descendants of Benjamin Allen of New Brunswick, Canada, into families. I also needed to prove their relationships to each other. We found hundreds of descendants, all surnamed Allen, concentrated in a single county. Some Allens in the area were not descendants of Benjamin Allen. We could not be sure that we had the Allen families linked together correctly until we linked together the information about each individual. Confusion reigned because many had the same name and similar birthdates. Some used a middle name to avoid confusion with a neighbor down the street. Some even had two middle names. A spouse who used more than one given name became another problem.

I decided to use the spreadsheet to display a time line for each individual. The individuals of interest were listed in successive Canadian census records (1851, 1861, 1871, 1881, and 1891). When I knew dates of birth, death, and marriage and names of spouses of Allens, I entered them in the timeline. A location was linked to age in each census to track the individual using the letter code with the age. I did not always have an accurate birth date, because of calculating the age from the census records, but a best estimate of the birth date was always included in the spreadsheet.

A row is used for each individual on the spreadsheet. For the timeline, there is a column for each source type: census, birth, death date, marriage date, etc. Rows of family members are kept

together if they lived at the same location. This is accomplished by placing them in consecutive rows and placing a box around the ages of the individuals who appeared in each census at a single residence. If the individual left home, he appears on a different line, but the census appearances are cross-referenced by indicating other line numbers where that individual appears. Usually two lines are needed for each individual, in one where he appears as a child with his parents and in the other as a parent with a family. A unique reference number is included to identify each line.

An individual might appear with age 25 in the 1851 census and ages 35, 45, and 55 in later censuses. Absence of a name on a census may be an indication of a move west or of a death. A move away from an area eliminated a family from confusion with others of the same name, so data from far away were not added to the spreadsheet. The missing name could also mean a marriage (and name change) for a female. The linking of families is a key to determining moves from the area, as families tended to move together. If a child did not appear in a family, it was usually an indication of a death. This is a portion of the spreadsheet:

ref #	surname	given	Census: 1851	1861	1871	1881	1891	1901	birth	mg.	spouse	death
17	Allen	Jonah	bot 30	bot 40	bot 50	bot 60			1821	1846	Ellen Read	
18	Read	Ellen or Eleanor	23	34	44	54			1826	1846	Jonah Allen	
19	Allen	Eliza Jane	4	14					1847			
20	Allen	Charles Perinton	2	12					1849			
21	Allen	Sarah Amelia	x	10	20	M	M		1851	1871	Howard Murray	
22	Allen	Jacob Judson	x	8	18	28			1852			
22.1	Allen	Joseph Crandall	x	6		n71	n71		1854			
23	Allen	Joshua Spurgeon	x	4	13	23			1857			
24	Allen	Arthur B.	x	x	7	17			1863			
25	Allen	Wilbur F. or William	x	x	5	15			1865			
26	Allen	Stephen Trenholm	x	n59	N59	bot 27	n736		1853	1876	Sarah E. Wall	
27	Wall	Sarah E.	unm	unm	unm	30	D		1850	1876	Stephen T. Allen	1877
28	Allen	Ephraim	N216		bot 48	bot 58	D		1823	1852	Elizabeth W. Allen	1894
29	Allen	Elizabeth W.	n54???		37	47	bot 56		1834	1852	Ephraim Allen	1908
30	Allen	Sarah			31				1839			
31	Allen	Nelson Watson	x		17	27	36		1854	1875	Emma Spence	
32	Allen	Major	x	x	x		15		1876		domestic	
33	Allen	Strang Sylvester	N500	n500	bot 32	?	bot 50	bot 60	1838	1866	Caroline Allen	1918
33.1	Allen	Caroline	N551	n551	23	32	43	bot 53	1848	1866	Strang Sylvester	1911
34	Allen	Willie E.	x	x	1	12	21		1869			
35	Allen	Rozala Lenie	x	x	x	9	19		1871			
36	Allen	Hanson	x	x	x	4	14		1876			
37	Allen	Caroline	x	x	x	1	11		1879			
38	Allen	Major	x	x	x	4	14	bot 24	1876			
869	Allen	Oakley S.	x	x	x	x	8	bot 18	1882			
870	Allen	Gracie P.	x	x	x	x	6	bot 16	1884			

Recognizing the same family in two different census entries is quite easy. The spouses' names and children's names will match with few exceptions. One exception occurs when a spouse dies and there is a second marriage. Children from a wife's first marriage are often helpful clues to connect families where there are few other clues. Often a surviving spouse will go to live with one of the children, which is another useful clue that comes from the census.

2. Using the sort or search routines

The primary use of the spreadsheet is to connect the children as adults with their parents. Once all the Allens in the area were entered, the rows were numbered using a special column, so the data can be sorted and then returned to the original order. When I sorted the data by birth date, individuals appearing as an adult and as a child would be nearby. Where there are two or three Williams with similar birthdates, they may be identified by the date that they left their parents'

homes. Individuals who use different names will also appear close together after sorting by birth date. They could be correlated to match individuals who appeared in more than one row. By sorting by sex and then by birth date, I place candidates who appear to be matched as the same individual in close vertical proximity in the spreadsheet. Where there are several Johns or Williams with similar ages, their identities can sometimes be resolved by the patterns of the timeline. For example, if one of the Johns in question stayed at home and appeared in a census after the other John married, the question is resolved for us. If one of the Johns appeared consistently to be a year older than the other, we may have a clean separation of the two.

A benefit of sorting by birth date is to clear up confusion over names. For example, William Trenholm Allen was recorded as a child with the name William and as an adult as Trenholm Allen because there was another William nearby. With the spreadsheet sorted by birth date, we may see entries William T. and W. Trenholm as evidence that these are the same individual.

We can also link marriages of the females with the individual as a child. Early marriage data in New Brunswick does not name the parents or give the age of the bride. The record would usually indicate whether she was a widow. Where there are several that could be the bride or groom in a marriage, we can eliminate some candidates because they appear in the census as unmarried after the marriage date.

The spreadsheet is very helpful in working with others. Should a new marriage be found, sorting the data by first name and then date of birth can place all the candidates together. Candidates can be quickly eliminated because the spreadsheet summarizes the sources of information available for each individual. Should two researchers claim that an individual has two different sets of parents, one glance at the spreadsheet shows the feasibility of the claim.

Once two rows are found to refer to the same individual, the data are not merged into one row as that would destroy the family structure shown by the boxes. Where one row has data for an individual, the other rows for that same individual are filled in with the other row number, but only in the columns where the other row has data. It can then be seen at a glance from the blanks where the person is still not accounted for in the timeline.

I have found the spreadsheet a necessity for sorting and matching more than 800 names in the Allen families in and around Westmorland County, New Brunswick. It was too confusing just using the PAF program. I am sure the spreadsheet is not worth the trouble when the individuals are scattered geographically and can be identified by location, or in cases where the family is small enough that the confusion can be resolved byother means. I constructed the spreadsheet four years ago to solve the problems mentioned here, but I refer to it now to answer correspondence from others because I can quickly see the facts, sources, and relationships.