Identity in Census Records:
Using Biological and Cultural Identifiers to Link Census Records

AUTHOR: Beau Sharbrough

INTRODUCTION: There are a variety of identifiers available for those attempting to link recordsets of family history data. There are special considerations for attempting to link records in consecutive US census records. These considerations involve a combination of biological and cultural identifiers easily found in the US census beginning in 1850. Preliminary experiments indicate that the discriminatory power of this set of identifiers exceeds that of traditional methods such as name matching and soundex matching.

The US Census records are snapshots of families, often showing young parents with toddlers on one census and older parents with teenagers on the next. By matching the members of the family from one census to the next some links can be established even if some family members are not present, and if the records contain wildly variant names.

The author will demonstrate various linking approaches using:

Old Man and Old Woman (man and woman in same house, similar names in each recordset, similar birth years in each recordset). In the example below, a HOUSEKEY match gives three true matches and one false match.

1860 OldMan First Name	1860 Old Man Surname	1860 Old Man Age	1870 Old Man First Name	1870 Old Man Surname	1870 Old Man Age	1860 Old Woman First Name	1860 Old Woman Surname	1860 Old Woman Age	1870 Old Woman First Name	1870 Old Woman Surname	1870 Old Woman Age	HouseKey
John	LEATHERS	33	John V	LEATHERS	43	Emily	LEATHERS	29	Emily C	LEATHERS	39	mLJ365fLE366
George	LIPPS	32	Geo.	LRAISNIC?	45	Mary	LIPPS	30	Mary J	LRAISNIC?	40	mLG365fLM366
Jesse	KELSO	25	Jesse	KELSEY	35	Louisa	KELSO	20	Liza	KELSEY	30	mKJ367fKL368
J.	KENNOR	26	James W	KENNER	36	Margaret	KENNOR		24	Margaret C	KENNER	34	mKJ366fKM367

Old Woman and Oldest Boy (woman and oldest boy in same house, similar names in each recordset, similar birth years in each recordset)
Matching Children (three children in same house, similar names in each recordset, similar birth years in each recordset)

Various linking keys are demonstrated, including initials, soundex, coded names and variations on calculation of age such as Calculated Birth Year / 5.

Using this method, the author was able to track a family from Arkansas 1860, Arkansas 1870, and Texas 1880, even though the father was not in the house in 1880, and the mother's first name was alternately listed as Parasaid, Hamzade, and Parazade.

In some cases, it is easier to link a house than an individual. One characteristic of families – living in the same house – make it possible to identify patterns in US census data with great accuracy than tracking individuals. This paper describes methods of defining a fingerprint for a house or family based on the people who live there.

REFERENCES.
Genealogical Demography, edited by Bennett Dyke and Warren T Merrill, Academic Press 1980
Reconstructing Historical Communities, Alan Macfarlane, Cambridge Univ Press 1977
Handbook of Record Linkage, Howard B Newcombe, Oxford Univ Press 1988
US Census, 1860 and 1870.