

**USING A (LINEAGE LINKED) FAMILY PERSPECTIVE OVER HOUSEHOLD  
TO IMPROVE RECORD LINKAGE SUCCESS WITH CENSUS (AND OTHER)  
DATA COLLECTIONS**

**Family History Technology Workshop**

**by David S. Barss, AG<sup>®</sup> 28 Apr 2010**



  
**FAMILYSEARCH**  
WHERE GENERATIONS MEET

Our Proposal is that:

- Creating lineage linked families and pedigrees from census data will
- Provide a broader foot print
- Thereby increasing opportunities for matching
- And improving record linkage success

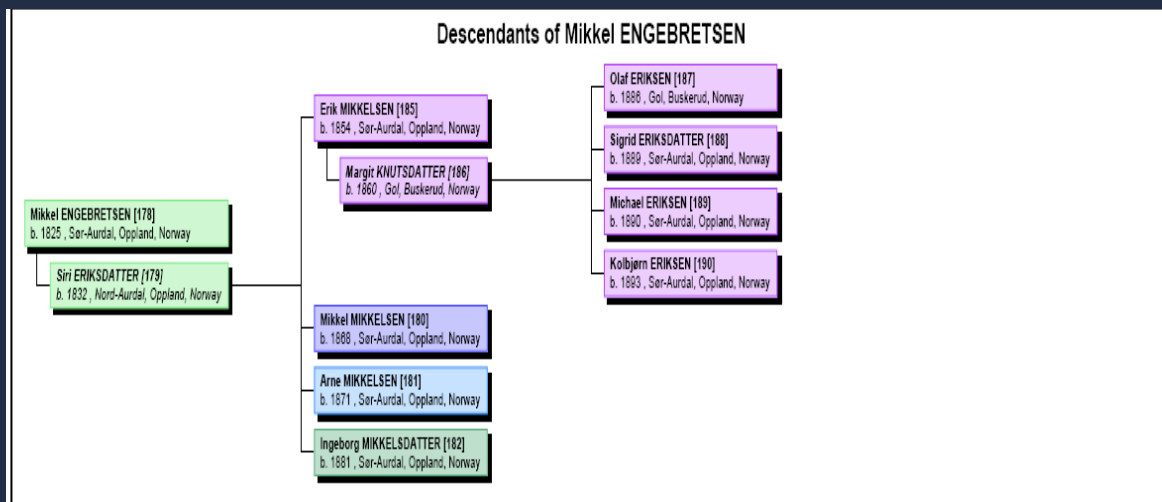
# Introduction

In other words - We think changing this...

	CENSUSNAME	RELATIONSHIP	SEX	BIRTHYR	AGE	MARST	RELATE	PERNOID	SPECOC	MOMLOC	POI
	Mikkel Engebretsen	Head	M	1825	075	M	0101	0001	0002		
	Siri Eriksdatter	Wife	F	1832	068	M	0201	0002	0001		
	Mikkel Mikkelsen	Son	M	1868	032	S	0301	0003		0002	0001
	Arne Mikkelsen	Son	M	1871	029	S	0301	0004		0002	0001
	Ingeborg Mikkelsdatter	Daughter	F	1881	019	S	0301	0005		0002	0001
►	Ole Dokken	Servant	M	1881	019	S	1211	0006			
	Olia Hansdatter	Servant	F	1864	036	W	1211	0007			
	Erik Mikkelsen	Son	M	1854	046	M	0301	0008	0009	0002	0001
	Margit Knutsdatter	Daughter-in-law	F	1860	040	M	0401	0009	0008		
	Olaf Eriksen	Grandson	M	1886	014	S	0901	0010		0009	0008
	Sigrid Eriksdatter	Granddaughter	F	1889	011	S	0901	0011		0009	0008
	Michael Eriksen	Grandson	M	1890	010	S	0901	0012		0009	0008

# Introduction

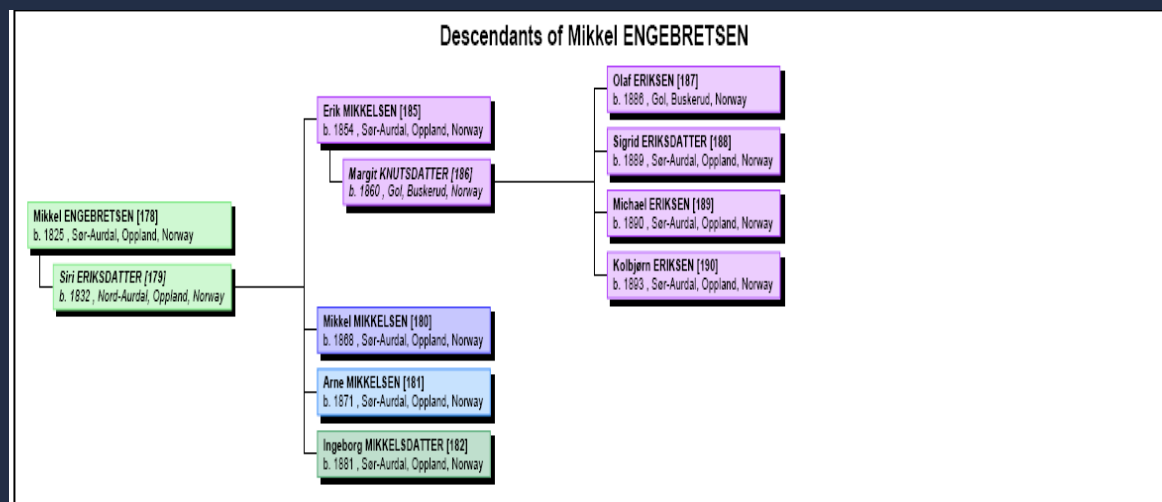
To this ...



# Introduction

Will improve record linkage success

	CENSUSNAME	RELATIONSHIP	SEX	BIRTHYR	AGE	MARST	RELATE	PERNUM	SPEC	WOMEOC	POP
	Mikkel Engebretsen	Head	M	1825	075	M	0101	0001	0002		
	Siri Eriksdatter	Wife	F	1832	068	M	0201	0002	0001		
	Mikkel Mikkelsen	Son	M	1868	032	S	0301	0003		0002	0001
	Arne Mikkelsen	Son	M	1871	029	S	0301	0004		0002	0001
	Ingeborg Mikkelsdatter	Daughter	F	1881	019	S	0301	0005		0002	0001
►	Ole Dokken	Servant	M	1881	019	S	1211	0006			
	Olaf Hansdatter	Servant	F	1864	036	W	1211	0007			
	Erik Mikkelsen	Son	M	1854	046	M	0301	0008	0009	0002	0001
	Margit Knutsdatter	Daughter-in-law	F	1860	040	M	0401	0009	0008		
	Olaf Eriksen	Grandson	M	1886	014	S	0901	0010		0009	0008
	Sigrid Eriksdatter	Granddaughter	F	1889	011	S	0901	0011		0009	0008
	Mikael Eriksen	Grandson	M	1888	016	S	0901	0012		0009	0008



# Introduction

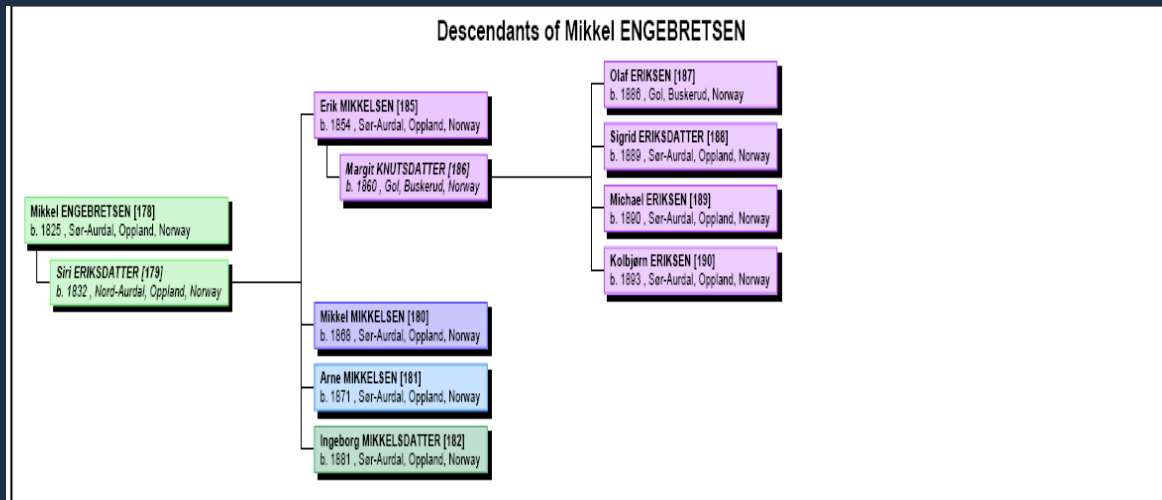
Traditionally work with census data:

- Uses the Household as the main unit of identification

	CENSUSNAME	RELATIONSHIP	SEX	BIRTHYR	AGE	MARST	RELATE	PERNO1	SPEC	MOME00	POP
	Mikkel Engebretsen	Head	M	1825	075	M	0101	0001	0002		
	Siri Eriksdatter	Wife	F	1832	068	M	0201	0002	0001		
	Mikkel Mikkelsen	Son	M	1868	032	S	0301	0003		0002	0001
	Arne Mikkelsen	Son	M	1871	029	S	0301	0004		0002	0001
	Ingeborg Mikkelsdatter	Daughter	F	1881	019	S	0301	0005		0002	0001
►	Ole Dokken	Servant	M	1881	019	S	1211	0006			
	Olja Hansdatter	Servant	F	1864	036	W	1211	0007			
	Erik Mikkelsen	Son	M	1854	046	M	0301	0008	0009	0002	0001
	Margit Knutsdatter	Daughter-in-law	F	1860	040	M	0401	0009	0008		
	Olaf Eriksen	Grandson	M	1886	014	S	0901	0010		0009	0008
	Sigrid Eriksdatter	Granddaughter	F	1889	011	S	0901	0011		0009	0008
	Mikkel Eriksen	Grandson	M	1890	010	S	0901	0012		0009	0008

# Introduction

- By using “the lineage linked family” as the main unit of identification



- We can accumulate more data from the record which improves opportunities for record linkage

Using a lineage linked family perspective:

- Preserves linkage that is already in the data
- Can provide a more accurate representation of the family
- Preserves families not related to the head of the household, such as servants or boarders, etc.



Adding genealogical interpretation and a few derived records  
allows us to capture:

- Sibling families
- In-law families
- Multi-generational families

## Introduction

Using local customs like patronymic naming patterns:

- Can provide a father's given name that is not stated in the census

# Introduction

If a marriage date is present we can:

- Identify and preserve “hidden families”
  - Children by the husband’s previous marriage
  - Children by the previous marriage of both spouses (yours, mine, and ours families)
  - A Mother that has remarried

# Census Data to Lineage Linked Family

- We have adopted the coding data fields and values used by the North Atlantic Population Project (NAPP)
- At the Minnesota Population Center
- University of Minnesota, Minneapolis

**NAPP**  
North Atlantic Population Project

MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA

Home Variables Create an Extract FAQ Contact Us

**The Project**

What is NAPP?  
Progress Report  
Release Dates  
Revision History  
FAQ

**Data**

NAPP Registration  
Create Data Extract  
Download Data Extract  
Analyze Data Online  
Citation and Use

**Documentation**

Samples  
Variables  
Census Questionnaires  
Other Resources

**Contact Us**

Feedback  
NAPP Participants

**Welcome to NAPP**

**NAPP News**

New! 2% sample of the 1851 Census of Great Britain  
Coming soon: Norway linked representative samples

**Other MPC Projects**

IPUMS-International  
IPUMS-USA  
IPUMS-CPS  
IHIS  
NHGIS

**IPUMS linked representative samples:** The IPUMS database links cases from the 1880 U.S. census to 1% samples of all other U.S. censuses from 1850 to 1930.

Copyright © Minnesota Population Center, University of Minnesota  
Funding provided by:

**NICHD**  
National Institute of Child Health and Human Development

**National Science Foundation**

# Census Data to Lineage Linked Family

Their coding process identifies:

- Each household
- The head of each household
- The relationship of each person to the head
- The position of each person within the household
- The position of each person's spouse (in house)
- The position of each person's father (in house)
- The position of each person's mother (in house)

# Census Data to Lineage Linked Family

We added a few more values to:

- Better handle Sibling-in-law relationships
  - Sibling-in-law = Your Spouse's Sibling
  - Sibling-in-law = Your Sibling's Spouse
  - Sibling-in-law = Your Spouse's Sibling's Spouse

## Census Data to Lineage Linked Family

We also added for Norway, a data field that:

- Allowed us to derive a father's given name
- From the child's patronymic surname
- When no father was present in the census

## Data Converter -CensusToGed

- These relationship and position pointers allow us to create lineage linked families

We worked with one of our vendors

- Pleiades Software Development , Inc.
- To create a tool that would convert a flat file with these relationships and position pointers to aGedcomFile for us.



# Data Converter -CensusToGed

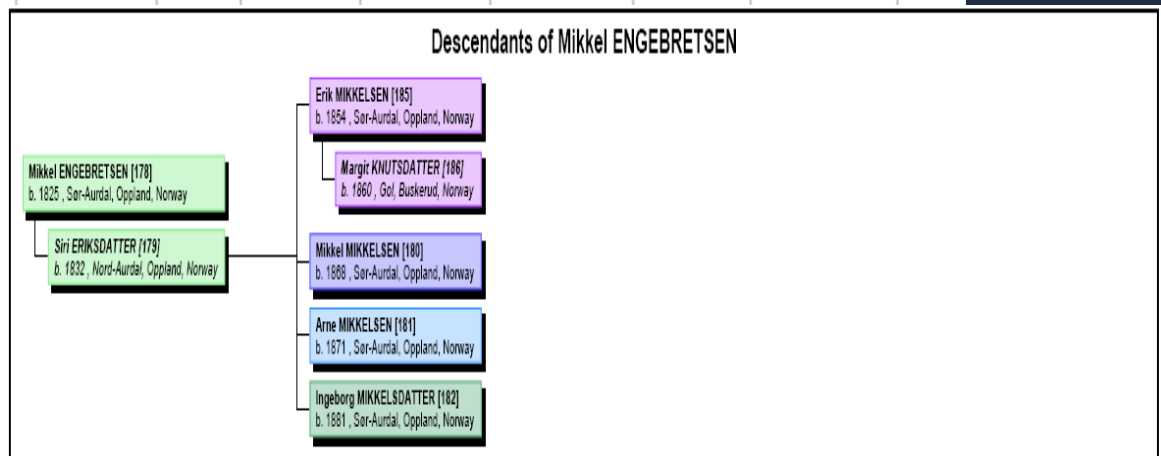
Using Relationship pointers alone - converter links:

- Head
- Spouse
- Child
- Sibling
- Parent
- Parent-in-law
- Sibling-in-law

# Data Converter -CensusToGed

- Relationships only – links simple families

	CENSUSNAME	RELATIONSHIP	SEX	BIRTHYR	AGE	MARST	RELATE	PERNOID	SPECOC	MOMNOID	POI
	Mikkel Engebretsen	Head	M	1825	075	M	0101	0001	0002		
	Siri Eriksdatter	Wife	F	1832	068	M	0201	0002	0001		
	Mikkel Mikkelsen	Son	M	1868	032	S	0301	0003		0002	0001
	Arne Mikkelsen	Son	M	1871	029	S	0301	0004		0002	0001
	Ingeborg Mikkelsdatter	Daughter	F	1881	019	S	0301	0005		0002	0001
►	Ole Dokken	Servant	M	1881	019	S	1211	0006			
	Olia Hansdatter	Servant	F	1864	036	W	1211	0007			
	Erik Mikkelsen	Son	M	1854	046	M	0301	0008	0009	0002	0001
	Margit Knutsdatter	Daughter-in-law	F	1860	040	M	0401	0009	0008		
	Olaf Eriksen	Grandson	M								
	Sigrid Eriksdatter	Granddaughter	F								
	Mikael Eriksen	Grandson	M								

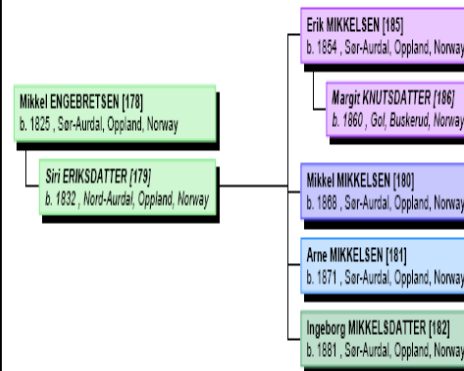


# Data Converter -CensusToGed

- Relationships only – but it misses some links

	CENSUSNAME	RELATIONSHIP	SEX	BIRTHYR	AGE	MARST	RELATE	PERNO1	SPECOC	MOMLOC	PO1
	Mikkel Engebretsen	Head	M	1825	075	M	0101	0001	0002		
	Siri Eriksdatter	Wife	F	1832	068	M	0201	0002	0001		
	Mikkel Mikkelsen	Son	M	1868	032	S	0301	0003		0002	0001
	Arne Mikkelsen	Son	M	1871	029	S	0301	0004		0002	0001
	Ingeborg Mikkelsdatter	Daughter	F	1881	019	S	0301	0005		0002	0001
▶	Ole Dokken	Servant	M	1881	019	S	1211	0006			
	Olia Hansdatter	Servant	F	1864	036	W	1211	0007			
	Erik Mikkelsen	Son	M	1854	046	M	0301	0008	0009	0002	0001
	Margit Knutsdatter	Daughter-in-law	F	1860	040	M	0401	0009	0008		
	Olaf Eriksen	Grandson	M								
	Sigrid Eriksdatter	Granddaughter	F								
	Mikael Eriksen	Grandson	M								

Descendants of Mikkel ENGBRETSSEN



## Data Converter -CensusToGed

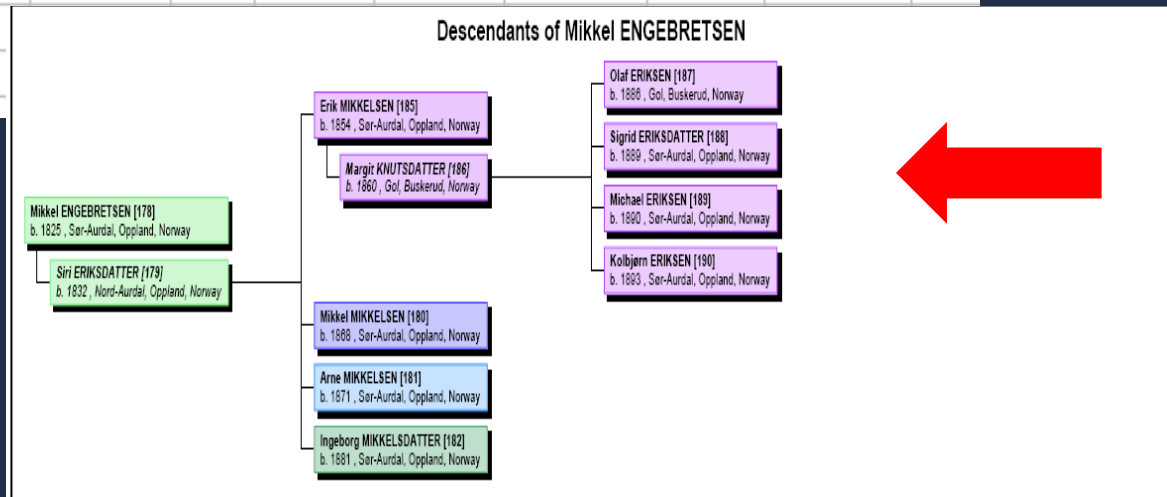
Using relationship and position pointers it links:

- All of the previously stated relationships - plus
- Grandchildren
- Hidden families
- Yours, Mine, and Ours families
- Boarder or Servant families
- And many others

# Data Converter -CensusToGed

- Using Relationships and Position Pointers

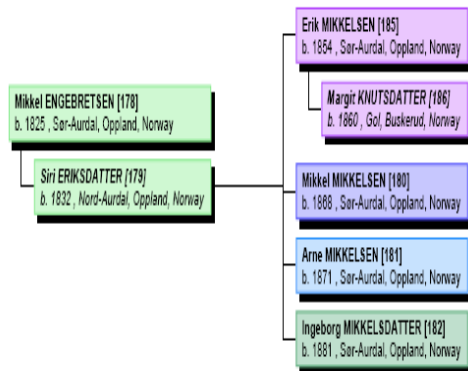
	CENSUSNAME	RELATIONSHIP	SEX	BIRTHYR	AGE	MARST	RELATE	PERNO1	SPECOC	MOMLOC	POI
	Mikkel Engebretsen	Head	M	1825	075	M	0101	0001	0002		
	Siri Eriksdatter	Wife	F	1832	068	M	0201	0002	0001		
	Mikkel Mikkelsen	Son	M	1868	032	S	0301	0003		0002	0001
	Arne Mikkelsen	Son	M	1871	029	S	0301	0004		0002	0001
	Ingeborg Mikkelsdatter	Daughter	F	1881	019	S	0301	0005		0002	0001
▶	Ole Dokken	Servant	M	1881	019	S	1211	0006			
	Olja Hansdatter	Servant	F	1864	036	W	1211	0007			
	Erik Mikkelsen	Son	M	1854	046	M	0301	0008	0009	0002	0001
	Margit Knutsdatter	Daughter-in-law	F	1860	040	M	0401	0009	0008		
	Olaf Eriksen	Grandson	M								
	Sigrid Eriksdatter	Granddaughter	F								
	Michael Eriksen	Grandson	M								



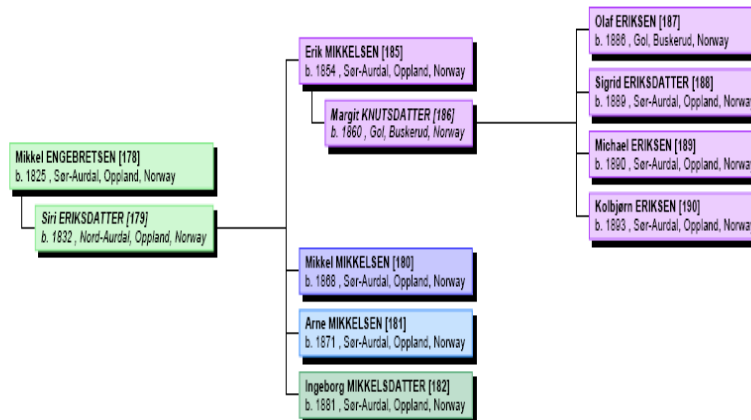
# Providing a Broader Foot Print

- By capturing all of the relationship data in the census as a lineage linked family
- We provide a broader foot print to match on

Descendants of Mikkel ENGBRETSSEN



Descendants of Mikkel ENGBRETSSEN



- We will come back to these grandchildren

## Providing a Broader Foot Print

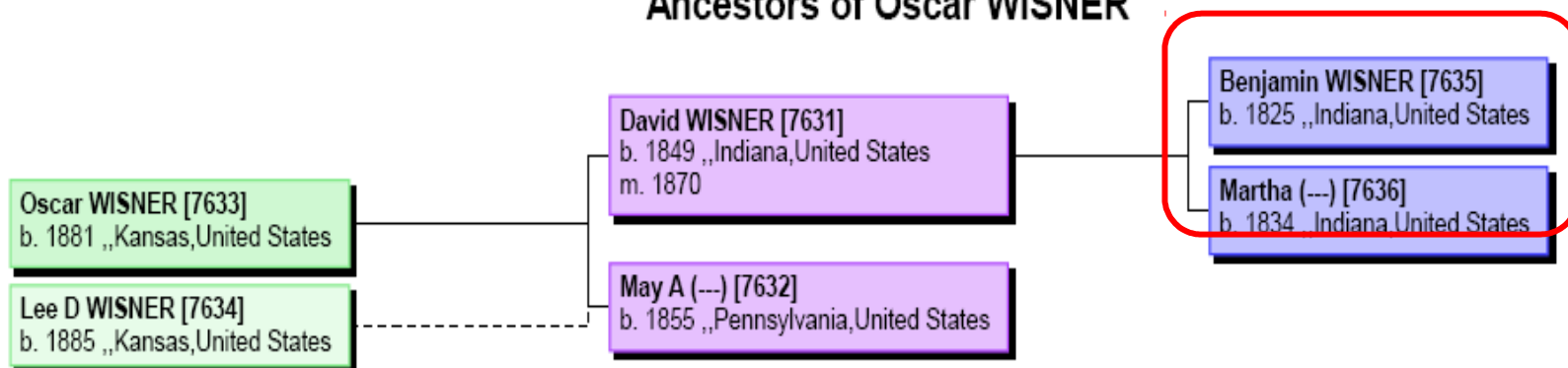
- More examples of how we have been able to **broaden the foot print** of the census are in the appendix to this paper
- I am going to quickly flash through most of them
- You can study them more closely later

# Household Including Parents

1900 US Census for Lewis County Washington

CENSUS_NAME	RELATIONSHIP	SEX	BIRTH_DATE	MSTS	MYRS	MDATE	MTHR	LVNG	RELATE	PERNUM	SPLOC	MOMLOC	POPLO
David Wisner	Head	M	Sep 1849	M	30	1870			0101	0001	0002	0006	0005
May A Wisner	Wife	F	Apr 1855	M	30	1870	6	4	0201	0002	0001		
Oscar Wisner	Son	M	May 1881	S					0301	0003		0002	0001
Lee D Wisner	Son	M	Aug 1885	S					0301	0004		0002	0001

## Ancestors of Oscar WISNER





# Household of siblings - DerivedFath.

1900 US Census for Lewis County Washington

CENSUS_NAME	RELATIONSHIP	SEX	BIRTH_DATE	MSTS	MYRS	MDATE	MTHR	LVNG	RELATE	PERNUM	SPLOC	MOMLOC	POPLOC
Charles C Henderson	Head	M	Oct 1872	S					0101	0001			
Joseph A Henderson	Brother	M	Jun 1876	S					0701	0002			
Albert E Henderson	Brother	M	Jan 1885	S					0701	0003			
Carl R Henderson	Brother	M	Feb 1887	S					0701	0004			
Flora E Henderson	Sister	F	Apr 1880	S					0701	0005			

## Ancestors of Charles C HENDERSON

Charles C HENDERSON [5787]  
b. 1872 „Massachusetts,United States

(---) HENDERSON [5793]  
b. „Massachusetts,United States

Joseph A HENDERSON [5788]  
b. 1876 „Massachusetts,United States

Flora E HENDERSON [5791]  
b. 1880 „Kansas,United States

Florence A HENDERSON [5792]  
b. 1882 „Kansas,United States

Albert E HENDERSON [5789]  
b. 1885 „Kansas,United States

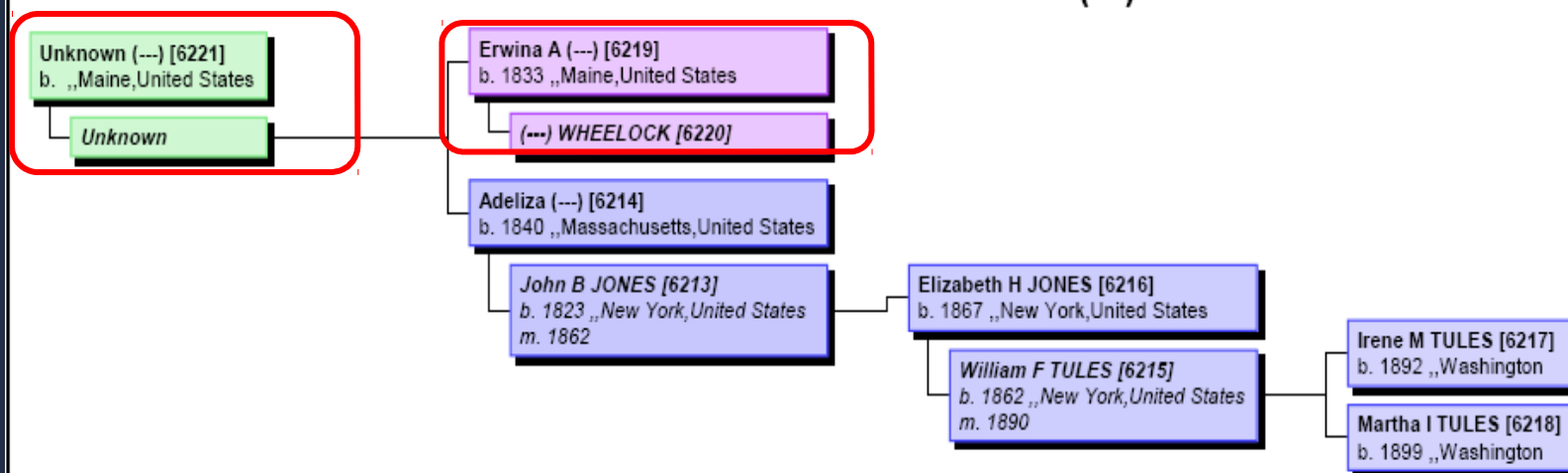
Carl R HENDERSON [5790]  
b. 1887 „Kansas,United States

# Sister-in-law (DerivedFath. & Hus)

1900 US Census for Lewis County Washington

CENSUS_NAME	RELATIONSHIP	SEX	BIRTH_DATE	MSTS	MYRS	MDATE	MTHR	LVNG	RELATE	PERNUM	SPLOC	MOMLOC	POPLOC
John B Jones	Head	M	Jan 1823	M	38	1862			0101	0001	0002		
Adeliza Jones	Wife	F	Jan 1840	M	38	1862	1	1	0201	0002	0001		
William F Tules	Son In Law	M	May 1862	M	10	1890			0401	0003	0004		
Elizabeth H Tules	Daughter	F	Dec 1867	M	10	1890	2	2	0301	0004	0003	0002	0001
Irene M Tules	Grand Daughter	F	Feb 1892	S					0901	0005		0004	0003

## Descendants of Unknown (---)

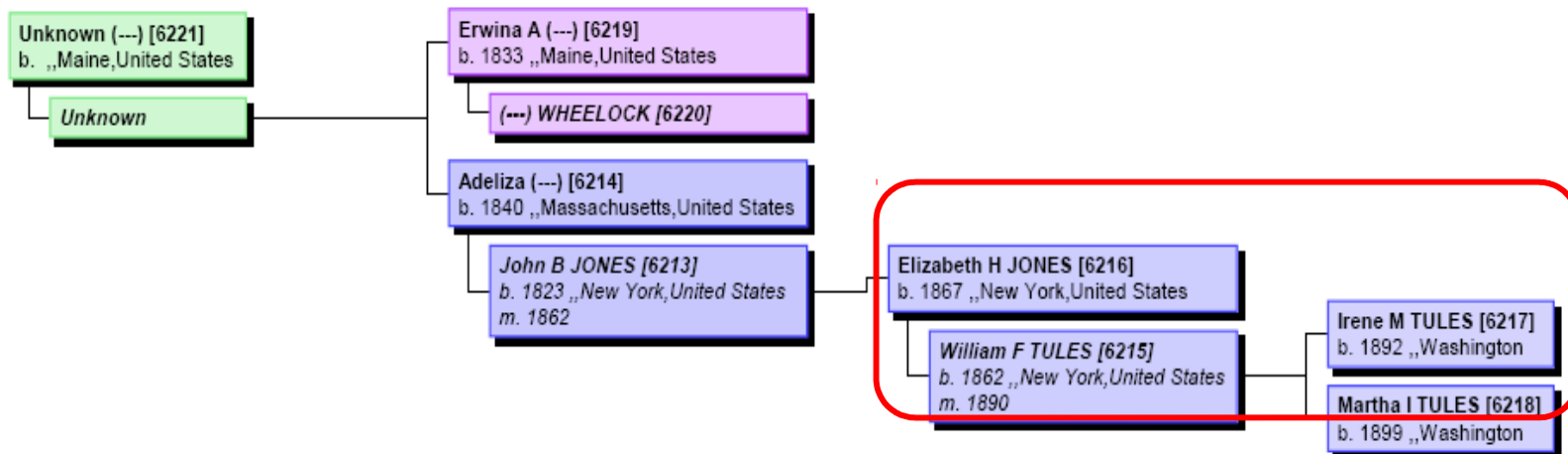


# Son-In-Law (with Family)

1900 US Census for Lewis County Washington

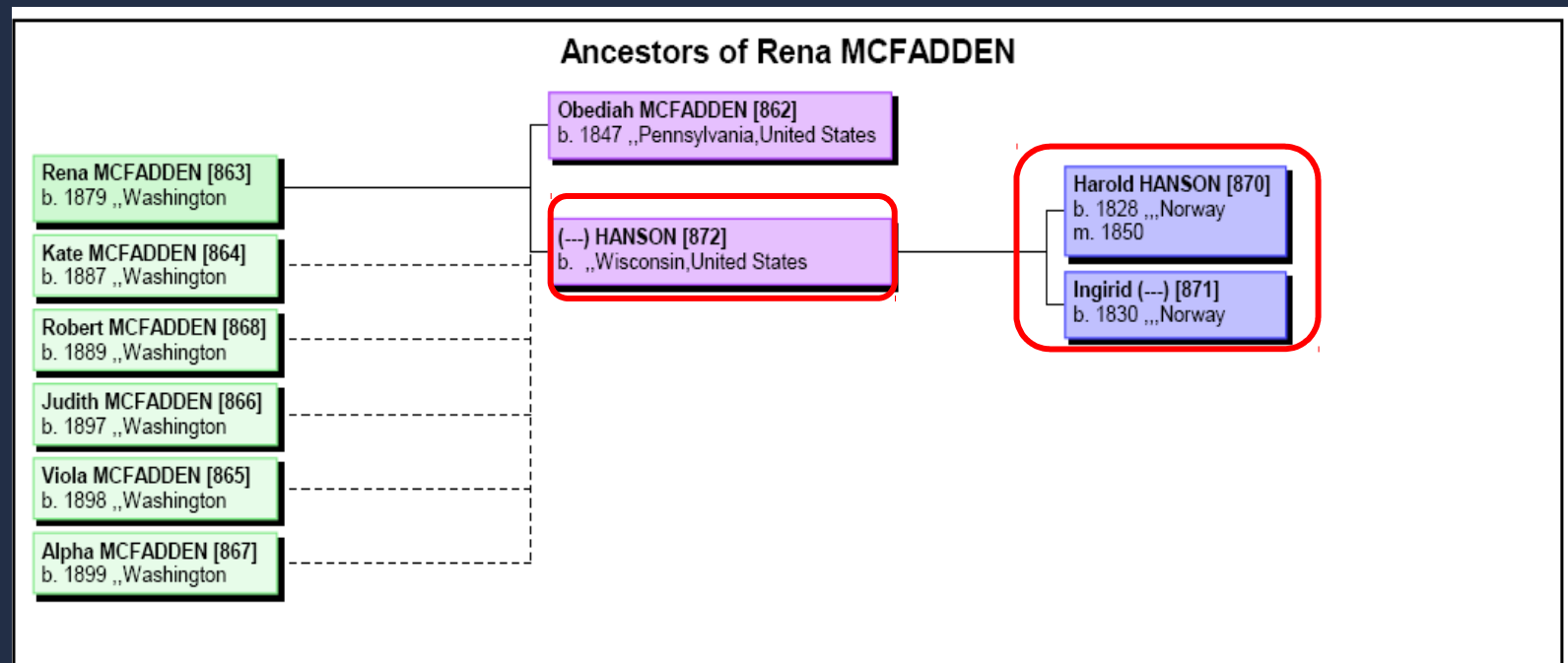
CENSUS_NAME	RELATIONSHIP	SEX	BIRTH_DATE	MSTS	MYRS	MDATE	MTHR	LVNG	RELATE	PERNUM	SPLOC	MOMLOC	POPLOC
John B Jones	Head	M	Jan 1823	M	38	1862			0101	0001	0002		
Adeliza Jones	Wife	F	Jan 1840	M	38	1862	1	1	0201	0002	0001		
William F Tules	Son In Law	M	May 1862	M	10	1890			0401	0003	0004		
Elizabeth H Tules	Daughter	F	Dec 1867	M	10	1890	2	2	0301	0004	0003	0002	0001
Irene M Tules	Grand Daughter	F	Feb 1892	S					0901	0005		0004	0003

## Descendants of Unknown (---)



# Parents-in-law (by Derived Spouse)

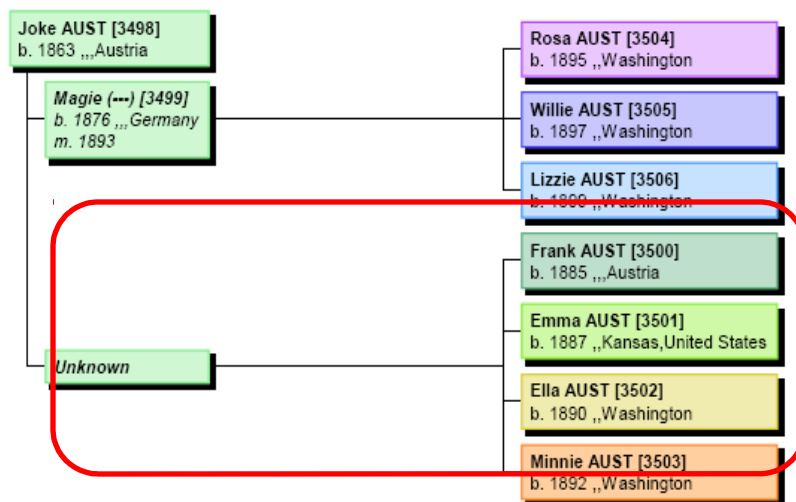
1900 US Census for Lewis County Washington													
CENSUS_NAME	RELATIONSHIP	SEX	BIRTH_DATE	MSTS	MYRS	MDATE	MTHR	LVNG	RELATE	PERNUM	SPLOC	MOMLOC	POPLOC
Obediah Mcfadden	Head	M	Dec 1847	W					0101	0001			
Rena Mcfadden	Daughter	F	Dec 1879	S					0301	0002			0001
Kate Mcfadden	Daughter	F	Apr 1887	S					0301	0003			0001
Viola Mcfadden	Daughter	F	Dec 1898	S					0301	0004			0001
Judith Mcfadden	Daughter	F	Feb 1897	S					0301	0005			0001
Alpha Mcfadden	Daughter	F	May 1899	S					0301	0006			0001
Robert Mcfadden	Son	M	May 1889	S					0301	0007			0001
Valentine Hall	Nephew	M	Mar 1887	S					1031	0008			
John H. Hall	Family Head	M	Apr 1828	M	50	1850			0601	0009	0010		



# Hidden Family: Previous Marriage

1900 US Census for Lewis County Washington													
CENSUS_NAME	RELATIONSHIP	SEX	BIRTH_DATE	MSTS	MYRS	MDATE	MTHR	LVNG	RELATE	PERNUM	SPLOC	MOMLOC	POPLOC
Joke Aust	Head	M	Mar 1863	M	7	1893			0101	0001	0002		
Magie Aust	Wife	F	Oct 1876	M	7	1893	3	3	0201	0002	0001		
Frank Aust	Son	M	Jan 1885	S					0301	0003		0002	0001
Emma Aust	Daughter	F	Jun 1887	S					0301	0004		0002	0001
Ella Aust	Daughter	F	Oct 1890	S					0301	0005		0002	0001
Minnie Aust	Daughter	F	Jun 1892	S					0301	0006		0002	0001
Rosa Aust	Daughter	F	Jan 1895	S					0301	0007		0002	0001
Willie Aust	Son	M	Mar 1897	S					0301	0008		0002	0001

## Descendants of Joke AUST

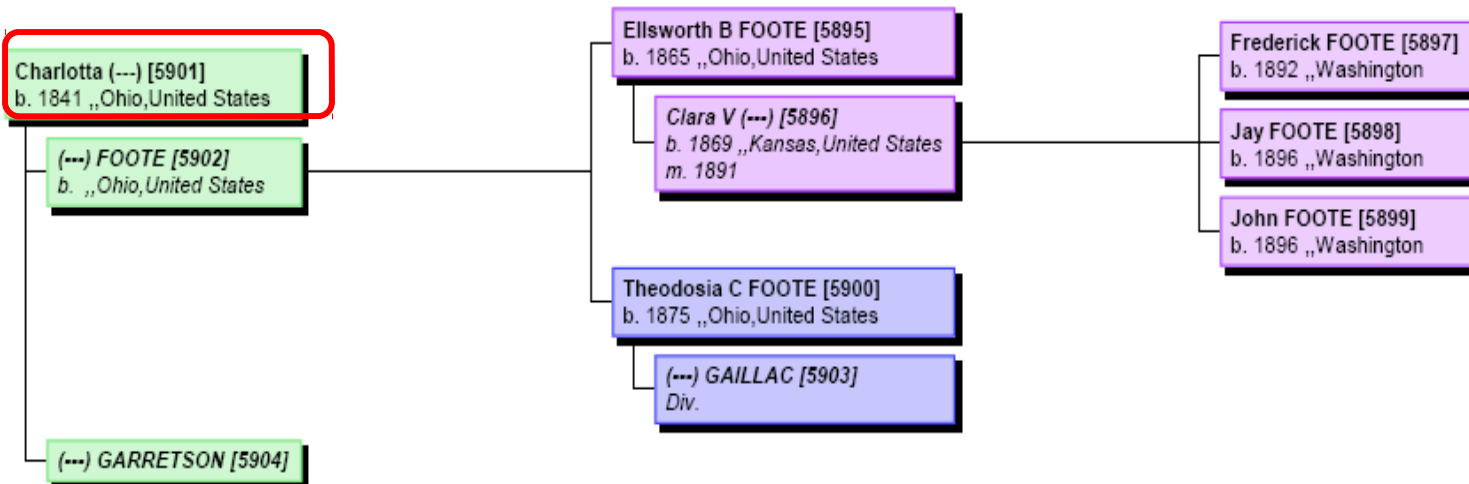


# Hidden Family: Mother remarried

1900 US Census for Lewis County Washington

CENSUS_NAME	RELATIONSHIP	SEX	BIRTH_DATE	MSTS	MYRS	MDATE	MTHR	LVNG	RELATE	PERNUM	SPLOC	MOMLOC	POPLOC
Ellsworth B Foote	Head	M	Jan 1865	M	9	1891			0101	0001	0002	0007	
Clara V Foote	Wife	F	Jan 1869	M	9	1891	3	3	0201	0002	0001		
Frederick Foote	Son	M	Apr 1892	S					0301	0003		0002	0001
Jay Foote	Son	M	May 1896	S					0301	0004		0002	0001
John Foote	Son	M	May 1896	S					0301	0005		0002	0001

## Descendants of Charlotta (---)

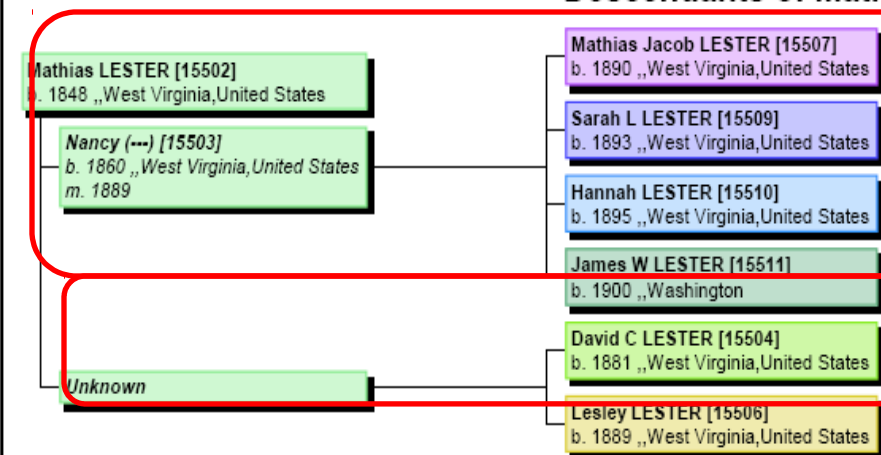


# Yours, Mine and Ours

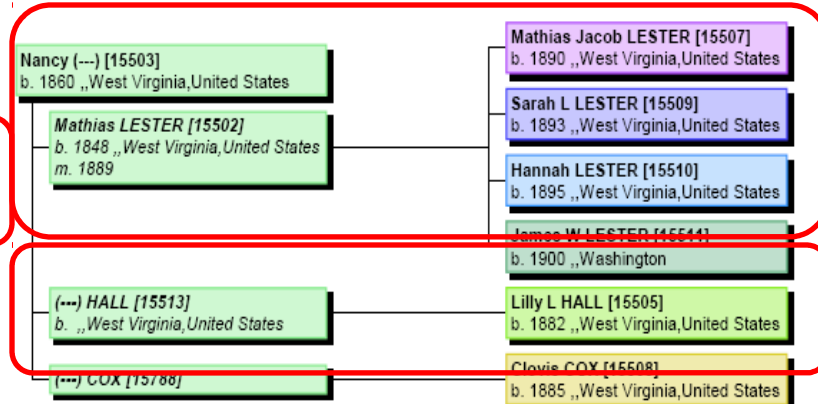
1900 US Census for Lewis County Washington

CENSUS_NAME	RELATIONSHIP	SEX	BIRTH_DATE	MSTS	MYRS	MDATE	MTHR	LVNG	RELATE	PERNUM	SPLOC	MOMLOC	POPLO
Mathias Lester	Head	M	May 1848	M	11	1889			0101	0001	0002		
Nancy Lester	Wife	F	Oct 1860	M	11	1889	9	7	0201	0002	0001		
David C Lester	Son	M	May 1881	S					0301	0003			0001
Lilly L Hall	Step Daughter	F	Dec 1882	S					0303	0004		0002	
Lesley Lester	Son	M	Mar 1889	S					0301	0005			0001
Mathias Jacob Lester	Son	M	May 1890	S					0301	0006		0002	0001
Clovie Cox	Step Son	M	Jun 1885	S					0303	0007		0002	
Sarah L Lester	Daughter	F	Apr 1893	S					0301	0008		0002	0001
Hannah Lester	Daughter	F	Feb 1895	S					0301	0009		0002	0001

## Descendants of Mathias LESTER



## Descendants of Nancy (---)



# Non-relatives (boarders or servants)

1900 US Census for Lewis County Washington

CENSUS_NAME	RELATIONSHIP	SEX	BIRTH_DATE	MSTS	MYRS	MDATE	MTHR	LVNG	RELATE	PERNUM	SPLOC	MOMLOC	POPLO
Mary A Reeder	Head	F	Jun 1850	W			5	5	0201	0001			
Joseph Reeder	Son	M	Jul 1882	S					0301	0002		0001	
Walter E Stanton	Boarder	M	Mar 1860	M	17	1883			1201	0003	0004		
Olive Stanton	Boarder	F	Jan 1864	M	17	1883	1	1	1201	0004	0003		

## Descendants of Mary A (---)

Mary A (---) [10496]  
b. 1850 „Missouri,United States

(---) REEDER [10502]  
b. „Ohio,United States

Joseph REEDER [10497]  
b. 1882 „Missouri,United States

## Descendants of Walter E STANTON

Walter E STANTON [10498]  
b. 1860 „Illinois,United States

Olive (---) [10499]  
b. 1864 „Iowa,United States  
m. 1883

Anna STANTON [10501]  
b. 1890 „,England

Fern STANTON [10500]  
b. 1896 „California,United States

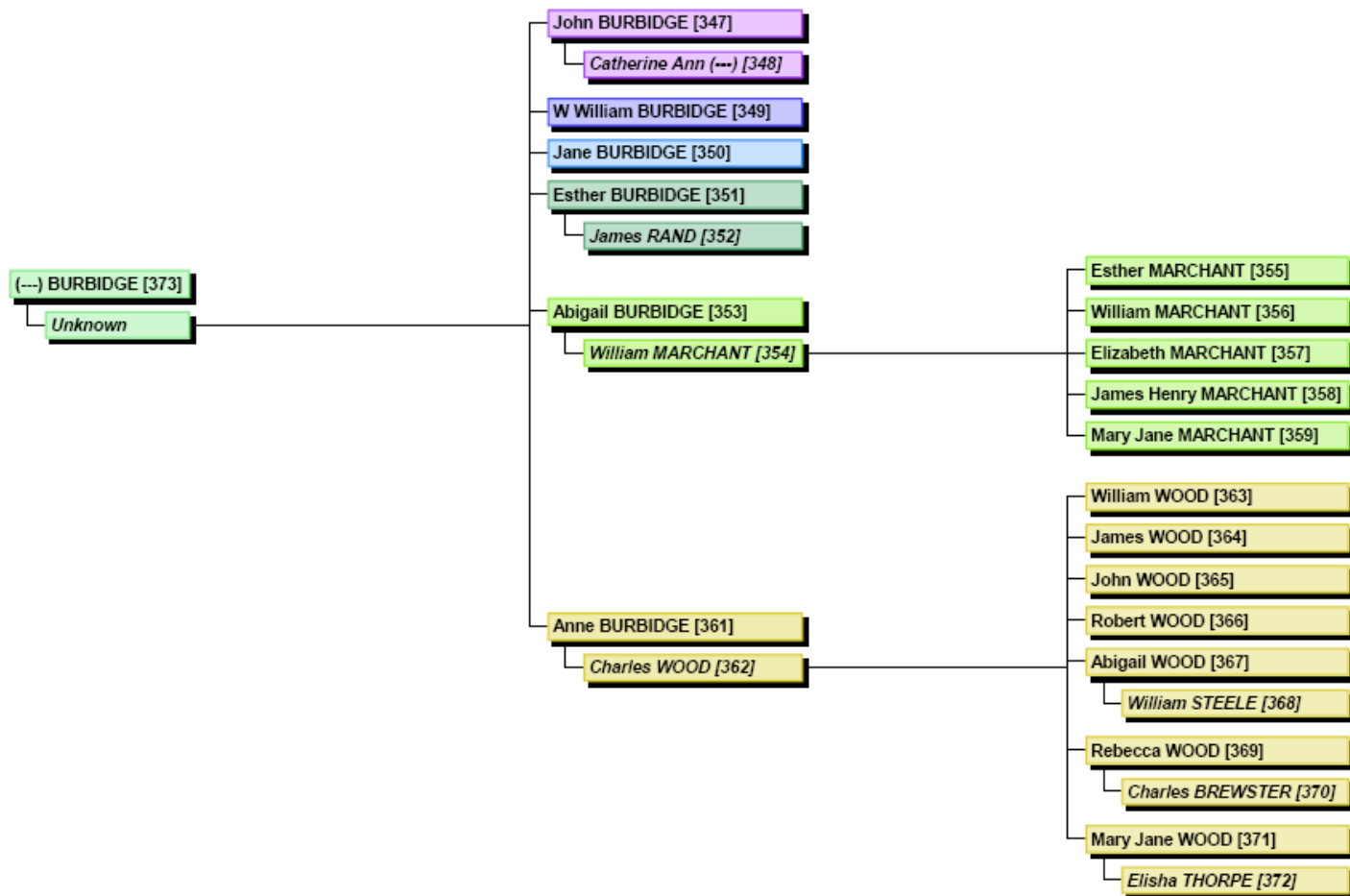


# Other Record Types - Probate

## King County Probate 1784 - 1949

ROLE	SURNAME	GIVENNAMES	MARRIEDSURNAME	SEXCODE	RELATIONSHIP	RELATE	PERNUM	SPLOC	MOMLOC	POPLOC
Principal	BURBIDGE	John		M	Head	0101	0001	0002		
Spouse		Catherine Ann	BURBIDGE	F	Wife	0201	0002	0001		

### Estate of John BURBIDGE (Probate 1853)



## Testing Record Linkage Success

- Now we need to answer the question is it worth the effort it takes to capture that lineage linked family data.

## Testing Record Linkage Success

To test our proposal we merged:

- Bygdebokand 1900 Norway Census data for Sør-Aurdal, Oppland, Norway
- Bygdeboksare local histories often containing extended family lineages
- Sør-Aurdalhad 4Bygdeboks
  - which we extracted using PAF and
  - merged into one lineage linked database

## Testing Record Linkage Success

We focused our attention on grandchildren:

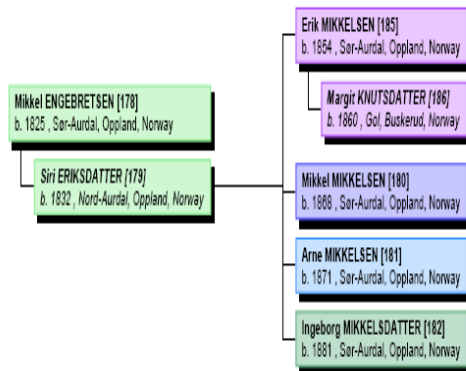
- There were 38 grandchildren in 1900Sør-Aurdal
- We found 24 of the grandchildren in theBygdebokdata
- From 10 households in the census
- These households had a connecting parent in the census data (giving at least a 3 generation link)

# Testing Record Linkage Success

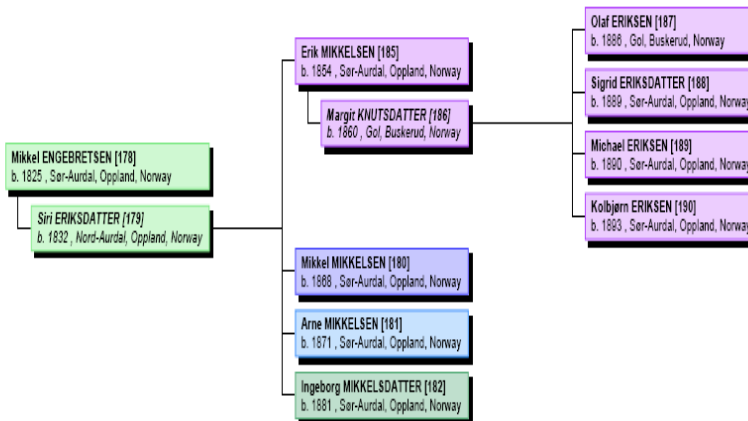
For our Census Data Samples we used:

- 1900 converted using pointers – Linked
- 1900 converted without pointers - Unlinked
- (grandchildren were not linked to family)

Descendants of Mikkel ENGE



Descendants of Mikkel ENGBRETSSEN



## Testing Record Linkage Success

- We used GenMergeDB as our merging tool
  - With the same settings in each test
- The Bygdebokdata also remained constant
- Therefore the only changing variable was the linked and unlinked nature of the census data

## Merging Results

- Bygdebokand 1900 Census with Links
  - 18 of 24 grandchildren were matched (75%)
- Bygdebokand 1900 Census without Links
  - 0 of 24 grandchildren were matched (0%)

## Merging Results

- Second Test – Used the census full data set
  - For the unlinked sample we removed all of the family links from the linkedGedcomfile – left just individuals
- Bygdebok and Census – Linked
  - Found 3414 people in 1659 clusters
- Bygdebok and Census – Unlinked
  - Found 98 people in 49 clusters
- Only 3% of the matches in the Linked sample were found in Unlinked sample



## Conclusion

- The test results for the grandchildren showed that the lineage linked nature of the data was critical to merging success
  - 75% Linked sample matching verses 0% Unlinked
- The merging with the full Census with the Bygdebok data did only slightly better
  - Only 3% of the matches found in the linked sample were found using the unlinked sample

## Conclusion

- Our conclusion is ....

That there is a lot of record linking advantage to be gained by converting census data (or other data sources) to lineage linked family data before merging

## Questions and Contact Information

- Are there any questions ?
- Contact information:
- David S. Barss
- E-Mail: [barssds@familysearch.org](mailto:barssds@familysearch.org)
- Come see use at the Family Reconstitution table in the FamilySearch booth

**Thank you**