

High-Level View of a Source-Centric Genealogical Model: *“The Model with Four Boxes”*

Randy Wilson

March 9, 2005

Necessary Elements

- Source Authority
- Artifact Archive
- Structured Data Archive
- Family Tree

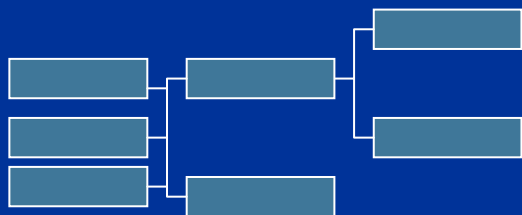


	- David	# 11 1	Son	1	
63 15	Gray, Carlo	# 7 8	Wife	1	Clergyman
	- Matilda	# 7 32	Daughter	1	Keeping House
	- Mary	# 7 12	Daughter	1	At Home
	- James	# 7 11	Son	1	At Home
	- George	# 7 9	Son	1	
	- Elizabeth	# 7 7	Daughter	1	
	- Martha	# 7 5	Daughter	1	
	- John	# 7 4	Son	1	
	- William	# 7 1	Son	1	

Household:

Name	Relation	Marital Status	Gender	Race	Age	Birthplace	Occupation	Father's Birthplace	Mother's Birthplace
Carlo GRAY	Self	M	Male	W	37	KY	Clergyman	KY	KY
Matilda GRAY	Wife	M	Female	W	32	VA	Keeping House	VA	VA
Mary GRAY	Dau	S	Female	W	12	VA	At Home	KY	VA
James GRAY	Son	S	Male	W	11	VA	At Home	KY	VA
George GRAY	Son	S	Male	W	9	VA		KY	VA
Elizabeth GRAY	Dau	W	Female	W	7	VA		KY	VA
Martha GRAY	Dau	S	Female	W	5	VA		KY	VA
John GRAY	Son	S	Male	W	4	VA		KY	VA
William GRAY	Son	S	Male	W	1	VA		KY	VA

Source Information:
 Census Place: Rose Hill, Lee, Virginia
 Family History Library Film: 1255375
 NA Film Number: T9-1375
 Page Number: 276C [View original image](#)
 Ancestry.com



1. Source Authority

List of all known potential sources of genealogical data.

- Assign unique ID to each source.
- Provide way to find existing sources
- Provide way to add new sources
- Assign unique id to each “page” of source

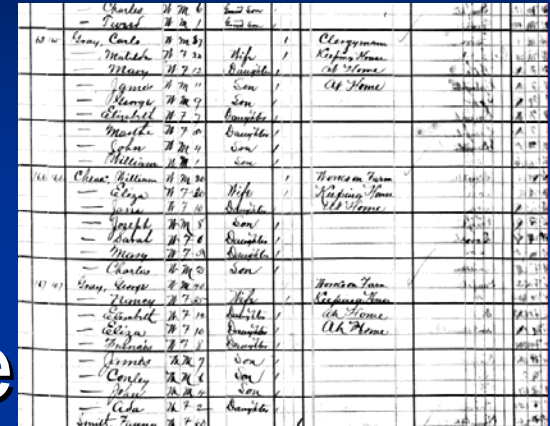


2. Artifact Archive

Hold scanned images for each page of each source

- Uses id of page and source from Source Authority.

- Enables indexing or deep extraction over the internet
- Enables verification and permanent preservation.



The image shows a scanned page from a handwritten ledger or index. The page is filled with rows of text, organized into columns. The text is written in cursive and includes names, dates, and other identifiers. The columns are separated by vertical lines, and the rows are separated by horizontal lines. The overall appearance is that of a historical record or a detailed index.

3. Structured Data Archive

a.k.a.: “Raw Data”, “Source Data”, “Evidence Database”, or “Extracted Records”

- Accurately represent what a source says
- Has unique ID for each “persona” (“name”/“reference to a person”) on each page.
- Contains names, dates, place, relationships, etc., that are clear from the source itself.
- Could possibly contain certain low-level assertions.

The screenshot shows a FamilySearch page for a household record from the 1880 United States Census. The page includes a navigation bar with 'HOME', 'SEARCH', 'SHARE', and 'LIBRARY' buttons. Below the navigation bar, there are links for 'Search for Ancestors', 'Research Guidance', 'Research Helps', 'Web Sites', 'Family History', and 'Library Catalog'. The main content area is titled 'Household Record' and includes a table of household members. The table has columns for Name, Relation, Marital Status, Gender, Race, Age, Birthplace, Occupation, Father's Birthplace, and Mother's Birthplace. Below the table, there is a 'Source Information' section with details about the census place, family history library film, NA film number, and page number. A 'View original image' button is also present.

Name	Relation	Marital Status	Gender	Race	Age	Birthplace	Occupation	Father's Birthplace	Mother's Birthplace
Cardo GRAY	Self	M	Male	W	37	KY	Clergyman	KY	KY
Matilda GRAY	Wife	M	Female	W	32	VA	Keeping House	VA	VA
Mary GRAY	Dau	S	Female	W	12	VA	At Home	KY	VA
James GRAY	Son	S	Male	W	11	VA	At Home	KY	VA
George GRAY	Son	S	Male	W	9	VA	At Home	KY	VA
Elizabeth GRAY	Dau	W	Female	W	7	VA		KY	VA
Martha GRAY	Dau	S	Female	W	5	VA		KY	VA
John GRAY	Son	S	Male	W	4	VA		KY	VA
William GRAY	Son	S	Male	W	1	VA		KY	VA

Source Information:
Census Place: Rose Hill, Lee, Virginia
Family History Library Film: 1255375
NA Film Number: T9-1375
Page Number: 276C

		Charles	W M 6	Grand Son	1				
		Tward	W M 1	Grand Son	1				
163	165	Gray, Carlo	W M 37		1	Clergyman			
		Matilda	W F 32	Wife	1	Keeping House			
		Mary	W F 12	Daughter	1	At Home			
		James	W M 11	Son	1	At Home			
		George	W M 9	Son	1				
		Elizabeth	W F 7	Daughter	1				
		Martha	W F 5	Daughter	1				
		John	W M 4	Son	1				
		William	W M 1	Son	1				
166	166	Cheas, William	W M 30		1	Work on Farm			
		Eliza	W F 20	Wife	1	Keeping House			
		Jane	W F 10	Daughter	1	At Home			
		Joseph	W M 8	Son	1				
		Sarah	W F 6	Daughter	1				
		Mary	W F 5	Daughter	1				
		Charles	W M 20	Son	1				
167	167	Gray, George	W M 40		1	Work on Farm			
		Nancy	W F 25	Wife	1	Keeping House			
		Elizabeth	W F 12	Daughter	1	At Home			
		Eliza	W F 10	Daughter	1	At Home			
		Francis	W F 8	Daughter	1				
		James	W M 7	Son	1				
		Conley	W M 6	Son	1				
		John	W M 4	Son	1				
		Ada	W F 2	Daughter	1				
		Smith, Fannie	W F 60		1				



HOME

SEARCH

SHARE

LIBRARY

Randy Wilson is signed on.

My Info | Sign Off

Search for
AncestorsResearch
GuidanceResearch
HelpsWeb
SitesFamily History
Library Catalog

Household Record

1880 United States Census

[Search results](#) | [Download](#)[Previous Household](#)[Next Household](#)

Household:

Name	Relation	Marital Status	Gender	Race	Age	Birthplace	Occupation	Father's Birthplace	Mother's Birthplace
Carlo GRAY	Self	M	Male	W	37	KY	Clergyman	KY	KY
Matilda GRAY	Wife	M	Female	W	32	VA	Keeping House	VA	VA
Mary GRAY	Dau	S	Female	W	12	VA	At Home	KY	VA
James GRAY	Son	S	Male	W	11	VA	At Home	KY	VA
George GRAY	Son	S	Male	W	9	VA		KY	VA
Elizabeth GRAY	Dau	W	Female	W	7	VA		KY	VA
Martha GRAY	Dau	S	Female	W	5	VA		KY	VA
John GRAY	Son	S	Male	W	4	VA		KY	VA
William GRAY	Son	S	Male	W	1	VA		KY	VA

Source Information:

Census Place Rose Hill, Lee, Virginia

Family History Library Film [1255375](#)

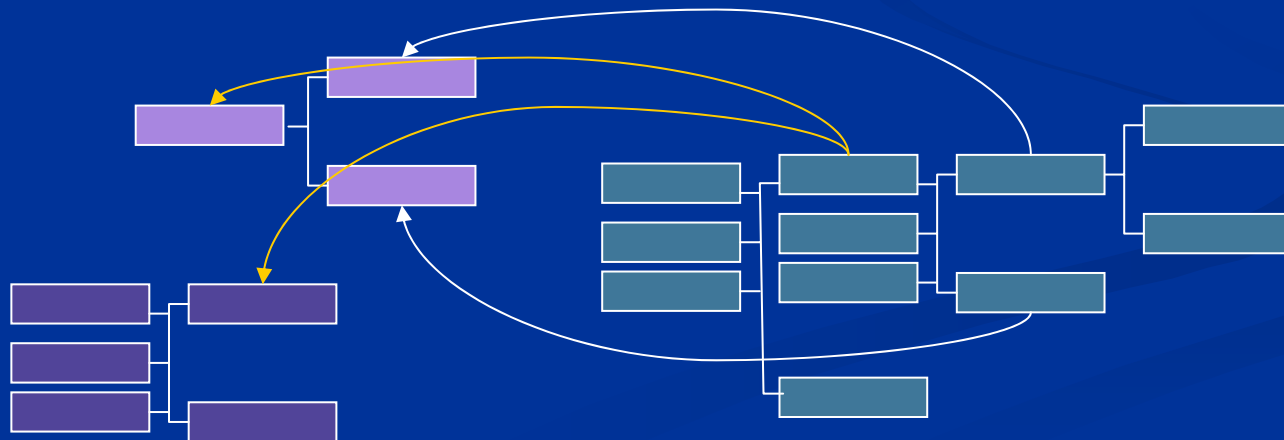
NA Film Number T9-1375

Page Number [276C](#)[View original image](#) at *Ancestry.com*

4. Family Tree

Represents our conclusions about who has lived and how they are related.

- Contains copies of information from the extraction archive, along with persona IDs that point to where the information came from.
- Links “personas” from different records together through relationships or merging/grouping.



“To Do” List

- ❑ **Locate** all sources in the world and add them to the Source Authority.
- ❑ **Extract** all genealogical data from all sources (usually from scanned images).
- ❑ **Link** all extracted personas into the Family Tree
- ❑ **Verify** that all of the above was done right
- ❑ **Perform** all ordinances.



Tasks for Users

- Enter what they know
- Extract/index data from images (perhaps in a locality of interest to them)
- Link extracted records into their line in the Family Tree
- Do verification work on extraction, linking.
- Take a name to the temple

Source Authority, *cont'd*

List of all sources, including, for example:

- Compiled family histories
- Records from courthouses and parishes
- Personal holdings (family bible, etc.)
- Cemeteries (pictures and/or transcriptions)
- Census records
- Memory of each user (i.e., personal knowledge)

Structured Data Archive

Important for:

- Computerized searching of data (traditional and record linkage)
- Pulling data into Family Tree for merging (grouping/linking).
- Knowing who on a page still needs work (and thus which sources still need work). Avoid missing people or repeating work forever.
- Simpler browsing of sources than using images.
- Additional context (e.g., who lives next door)

Extraction Archive, *cont'd*

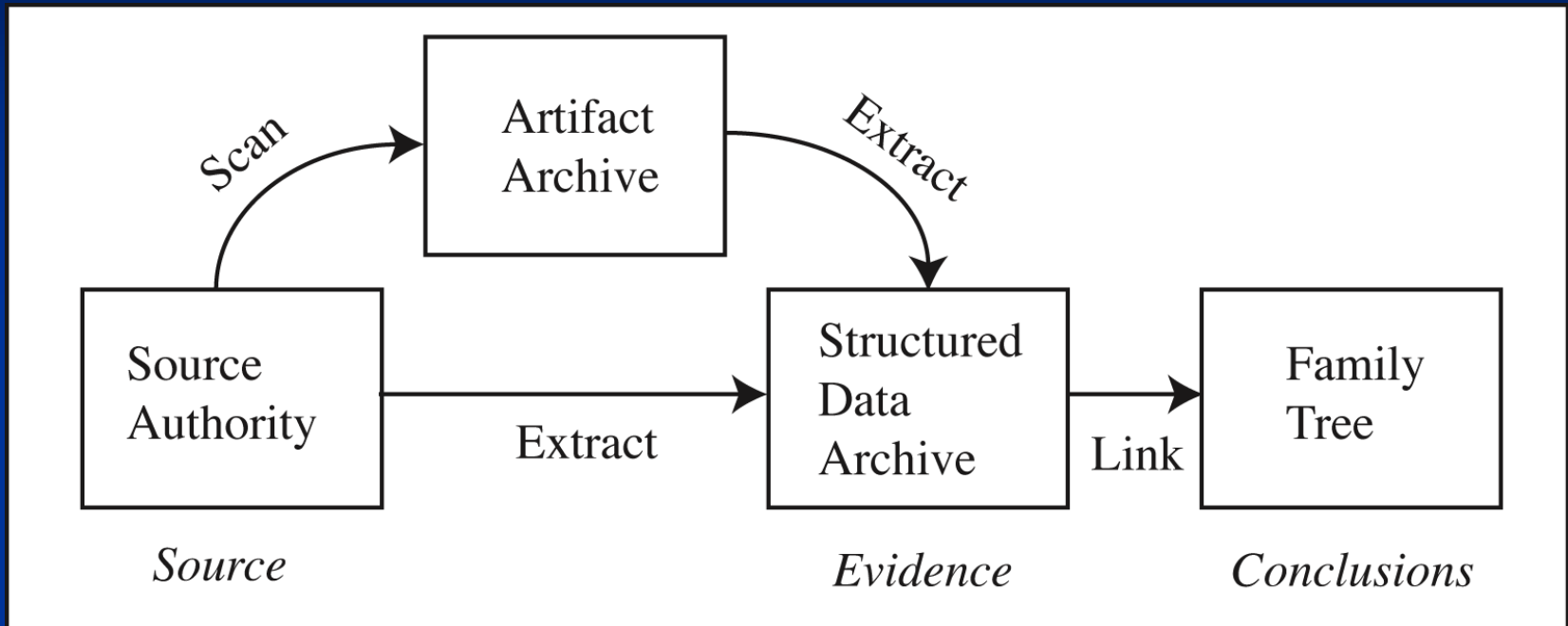
Also:

- Having intermediate step separates extraction work from linking/merging work and other conclusions.
- This makes it simpler to verify each step.
- It also makes it more clear where differences of opinion are coming from

Verification

- Currently thorough genealogists go back to the original source to confirm anything they find in an electronic database.
- This would take each person forever on the Family Tree.
- We must store the fact that each step has been verified so that eventually we can trust well-verified work and move on to something else.

“The Model with Four Boxes”



Source / Evidence / Conclusions

Record collections / extracted records / people

Summary

Source-centric approach allows completeness without endless duplication of effort.

Users can participate in various activities, but all of these can help move data from one point in the process to the next.

Industry needs to think in a source-centric way to enable true collaboration.

Questions? Ideas?

E-mail: wilsonr@ldschurch.org

Blogs:

eatslikeahuman.blogspot.com

source-centric-genealogy.blogspot.com