

## Who Is John Smith: Adventures In Genealogical Data Modeling

I'm no physicist, but they say when you look for an atom's component electrons, the electrons are both everywhere and nowhere at the same time. You can never actually see a single electron within the atom that houses the electron, yet everywhere you look, there is evidence of the electron in that exact point at that exact time.

In attempting to pinpoint where a person is within genealogical data, one runs into the same phenomenon. Consider the following. Here is some evidence that may be relevant to a fictional John Lester Smith. It is not meant to be an authoritative list but merely illustrates the current issue:

### Evidence

1. A birth certificate from Fayette County, Illinois, which states a male child named John Lester Smith was born at 401 South Third Street, Vandalia, Fayette County, Illinois, on 1 April, 1900 to parents John Smith and Mary Smith.
2. An interview with George Rogers Smith in which George refers to his father's brother, "Uncle John."
3. Marriage certificate from Fayette County, Illinois, of John Smith and Darcy Madsen who were married 5 April, 1921 in Fayette County, Illinois.
4. An entry in the 1910 U.S. Federal Census from Enumeration District 12, Ward 3, Vandalia, Fayette County, Illinois, showing in dwelling 14 household 16 at 401 3<sup>rd</sup> Street family members John Smith, aged, 32, and wife Darcy Smith, aged 30, with son Johnny Smith, aged 10, all born in Illinois.

Now, suppose I "know" there was a person in my family named John Lester Smith. How? Grandma told me. I heard about him all my life. This was a real person, and I identify him in my genealogy database. Maybe I even talked to him during a séance. (If so, it would make for an interesting source citation.) The point is, I have made an analysis and judgment that there is a person named John Lester Smith. I further think that the pieces of evidence above refer to this person I "know" existed, John Lester Smith.

What does the data show, independent of any analysis, that might be relevant to John Lester Smith?

1. A male named John Lester Smith was born at 401 South Third Street, Vandalia, Fayette County, Illinois, on 1 April, 1900 to parents John Smith and Mary Smith.
2. A male named John Smith distinct from John Lester Smith his son.
3. An Uncle John, brother of the father of George Rogers Smith.
4. A John Smith married Darcy Madsen on 5 April, 1921 in Fayette County, Illinois.
5. John Smith living in Vandalia in 1910 aged 32.
6. Johnny Smith, son of John and Mary Smith, aged 10, living with his mother and father in Vandalia.

OK, great. Now I want to put this data my computer's genealogy database.

Within my database, the identification of the person within a piece of evidence is done according to the name (or names) used in that evidence. Thus:

Person-name:  
Person-name-number=1  
Title=  
Firstname=John  
Middlename= Lester  
Lastname=Smith  
A. Referenced in a  
birth record

Person-name:  
Person-name-number=2  
Title=  
Firstname=John  
Middlename=  
Lastname=Smith  
A. Referenced in a  
birth record  
B. Referenced in a  
marriage record  
C. Referenced in a  
census record

Person-name:  
Person-name-number=3  
Title=Uncle  
Firstname=John  
Middlename=  
Lastname=  
A. Referenced in an  
interview

Person-name:  
Person-name-number=4  
Title=  
Firstname=Johnny  
Middlename=  
Lastname=Smith  
A. Referenced in a  
census record

But, where's the part that identifies John Lester Smith, the data representation of the whole person? That piece isn't here. To start describing this actual, real-life person, we need another object type entirely, an object which isn't always available in genealogy software as a separate entity:

```
real-person:  
  
PersonID=1  
  
Whole-person-name=  
  
Real-person-  
firstnameFirst=John  
  
Real-person-  
middlename=Lester  
  
Real-person-  
lastname=Smith  
  
Ref=person-name:1:A  
  
Ref=person-name:3:A  
  
Ref=person-name:4:A
```

Also, why are these references to the previous person-name records part of this real-person record? Where can I put that information into my database? Sources and source citations can be added to genealogy programs for the entries in the person-name records, but there is no way to document the evaluation of this information in today's software.

Now, you might say, "My database does that differently. It isn't a problem." Really? I challenge you to examine it carefully. I think you'll find that somewhere your evidence and your analysis of that evidence (in this case, expressed as your determination of evidence referring to a single person) get all mixed together in inappropriate ways.

I am not trying to debate the old question, "How do we know anything is real?" Rather, I want to point out that there is a layer of analysis missing in our current tools with which we attempt to document genealogical information, and absence of this analysis layer causes our data to be distorted. This isn't a particular problem for most cases, but when you get into a tricky case of an ancestor who has conflicting

evidence or other need for more detailed analysis, our ability for current genealogy databases to document these issues breaks down.

This document is far from complete, but it is meant to be a starting point for discussion. It needs further refinement to be sure. However, I hope it is useful in illustrating the need for a better, uniform analysis structure within genealogical software database applications.