

The background features a series of overlapping green triangles of various shades, creating a mountain-like silhouette. A wide, horizontal band of bright yellow with a fine, grainy texture runs across the middle of the image, separating the upper green section from the lower green section.

# Duplicate Removal

Family History Technology Conference Lightning Talk  
Dr. Mark Clement, Computer Science, Brigham Young University

# Duplicates are confusing



Family Tree Memories Search Indexing Temple

Messages Volunteer Get Help

Mark Joel Clement

Go to: [George Ware](#)



Person

[George Ware](#)  
LZSS-Z59

Events

birth 17 Jun 1737, Smarden, Kent, England  
death 9 Dec 1801, Lenham, Kent, England

Parents

■ Smith Ware  
● Elizabeth George

Spouse

● Elizabeth Ottaway

Possible Duplicates **1**

Merge by ID

Not a Match **0**

## Possible Duplicates for George Ware



Merging is a complex process in which you decide if two people are the same person. If they are, you choose which information should be kept. Please take the time necessary to carefully review each possible duplicate.

## Matching People (1 result)

Person

Events

Parents

Spouse



[George Ware](#)  
L4GV-G78  
●●●●○

● Elizabeth

[Review Merge](#)

Not a Match

# Duplicates and New Users

- Many users will give up when they are warned:
  - “Merging is a complex process in which you decide if two people are the same person. If they are, you choose which information should be kept. Please take the time necessary to carefully review each possible duplicate.”
- There are few hints provided about which individuals should be merged.
- This problem becomes more difficult as additional sources are added to familysearch



# A Modest Proposal

- Automatically search through source records to determine the frequency of the duplicate name and the likelihood that the individuals are duplicates.
- Provide the user with guidance as to which fields are likely to be correct in each of the duplicates. The presence of source documents and corroborating evidence could provide this likelihood
- Let the software determine the most likely resolution and allow the user to approve the merge.

# Using likelihoods

- Given original source data from census, marriage and death records, build the most likely tree automatically.
- Allow users to approve auto-built branches and incorporate them into the tree.