FAMILYSEARCH

DEALING WITH DUPLICATE RECORDS OF PEOPLE IN FAMILY TREE

A FAMILYSEARCH WHITE PAPER

21 JUNE 2012

EXECUTIVE SUMMARY

Both new.familysearch.org and Family Tree have duplicate records. Family Tree will provide a better solution for handling these duplicate records. It:

- Allows users to merge duplicate records, choosing which information to keep and which to archive.
- Fixes IOUS (Individuals of Unusual Size) records and prevents their creation.
- Allows users to correct records that were combined or merged inappropriately.
- Prevents the merging of wrong records with a new "not a match" feature.

INTRODUCTION

This whitepaper discusses how FamilySearch Family Tree will handle duplicate records of people.

Understanding how Family Tree provides a better solution requires first understanding:

- How new.familysearch.org handles duplication.

- How combined records are moved from new.familysearch.org to Family Tree.

DUPLICATES IN NEW.FAMILYSEARCH.ORG

New.familysearch.org identifies records that are "possible duplicates."

Users can either ignore these duplicates or combine them. When users combine records, they can choose to compare details of the two records or do a quick combine without looking at the details.

After duplicates are combined, the system does these things:

1. It creates a new record for the person. This is why the person identifier (PID) changes. This new record is often called the "combined record."
2. It puts the two duplicates into the newly created record.

3. It chooses default information (name, birth, etc.) for use on the new record's Summary tab.

Combining records often results in an issue that we call "IOUS." If a combined record acquires a large number of duplicates, it becomes an IOUS. The system has difficulty processing all of the information in an IOUS in a timely manner.

For this reason, new.familysearch.org limits the number of duplicates that can be combined. If combining two duplicates would create a record that is too big, the system does not let users combine them.

HOW COMBINED RECORDS MOVE FROM NEW.FAMILYSEARCH.ORG TO FAMILY TREE

The FamilySearch Family Tree is a replacement for new.familysearch.org. It will contain the pedigree for www.familysearch.org.

During the creation of Family Tree, the data is being moved from new.familysearch.org to the Family Tree.

Family Tree solves the IOUS issue because it does not keep all of the combined duplicate records. Instead, Family Tree brings over only some of the data.

For example, new.familysearch.org can have multiple versions of a person's name, gender, birth, christening, death, and burial information. Family Tree has only one version. When the information is transferred, Family Tree keeps the version that is selected on the person's Summary tab. The variations do not come over.

DUPLICATES IN FAMILY TREE

Family Tree will also have a possible duplicates feature, but it will work in a very different way.

Instead of combining duplicates, Family Tree helps users merge them, which means that users choose which data to keep and which to archive.

During a merge, Family Tree:

- Shows users all of the information and relationships from both duplicates.
- Lets users compare each piece of information and each relationship.
- Lets users choose which information to keep and which to archive. Because the Family Tree allows sources to be attached, users will have more information to use in making this decision.

After a merge, Family Tree:

- Keeps the merged record in the tree, leaving its person identifier unchanged.
- Adds to the merged record the information that was selected from the duplicate record.
- Archives the duplicate record, leaving its person identifier unchanged.
- Records these actions on the change history so that they can be undone if needed.
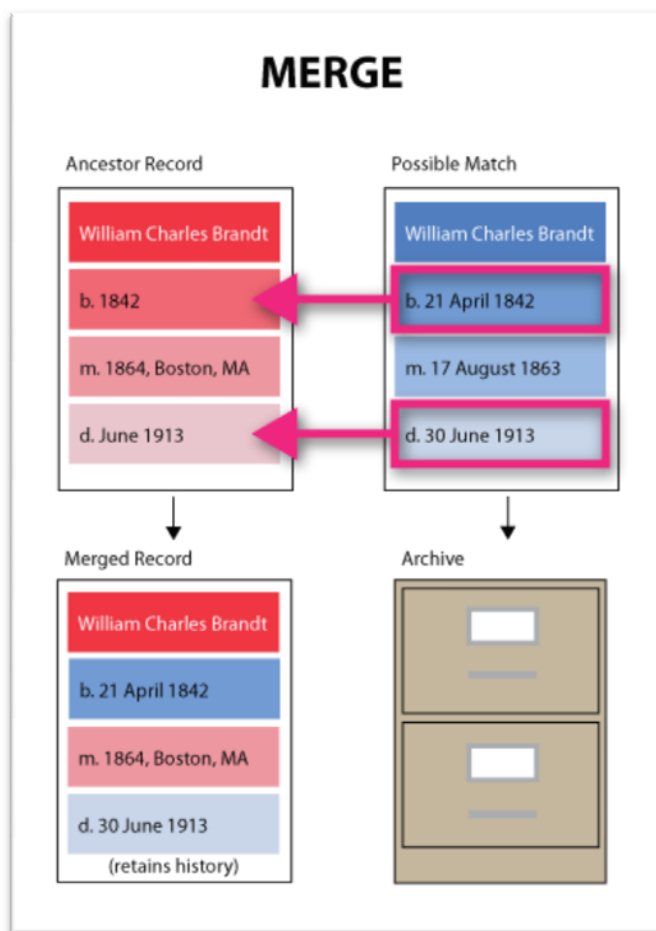
MERGE

Ancestor Record

William Charles Brandt

b. 1842

m. 1864, Boston, MA

d. June 1913

Possible Match

William Charles Brandt

b. 21 April 1842

m. 17 August 1863

d. 30 June 1913

Merged Record

William Charles Brandt

b. 21 April 1842

m. 1864, Boston, MA

d. 30 June 1913

(retains history)

Archive

Figure 1. How the merge feature will work in Family Tree.

INAPPROPRIATE MERGES

In new.familysearch.org, the records about different people are frequently combined inappropriately. Both new.familysearch.org and Family Tree provide tools to fix this problem. Family Tree's tools will be more powerful.

NEW.FAMILYEARCH.ORG

In new.familysearch.org, a user has to open the combined record and separate out the records that do not belong.

Separating records is not a simple or clean solution because of the way information is stored:

- Some pieces of information are stored only on the records that are contained within the combined record.
- Some pieces of information are stored only on the combined record. Some information can be in either place.

If someone changes a record after it is combined, the system stores the change on one of the involved records. The system can choose any of the records: the combined record or any of the records inside of it. There is no way to specify where the system should store the change.

Consequently, if a record is combined, and then someone changes it, the system chooses one of the records and stores the changed information on it. If all of the records are about the same person, this is not a problem.

However, if the records are about different people, the system might store the change on the record of the wrong person.

Later, when the records are separated, the change goes with the record that the system stored it on.

FAMILY TREE

Family Tree uses a person's change history to correct a bad merge. There will be two features used to do this:

- Undo the merge.
- Re-create the record that was merged and archived.

Undo the Merge (Unmerge)

If no one changes a person's record after a merge, a merge can be undone. Both the original and duplicate records are restored as if no merge had occurred.
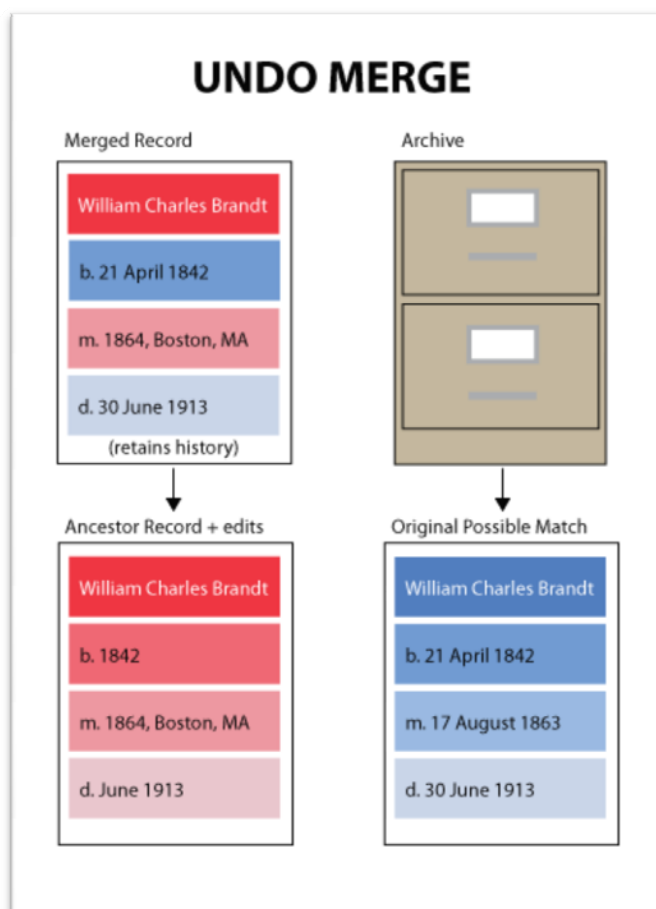
This graphic illustrates undoing a merge:

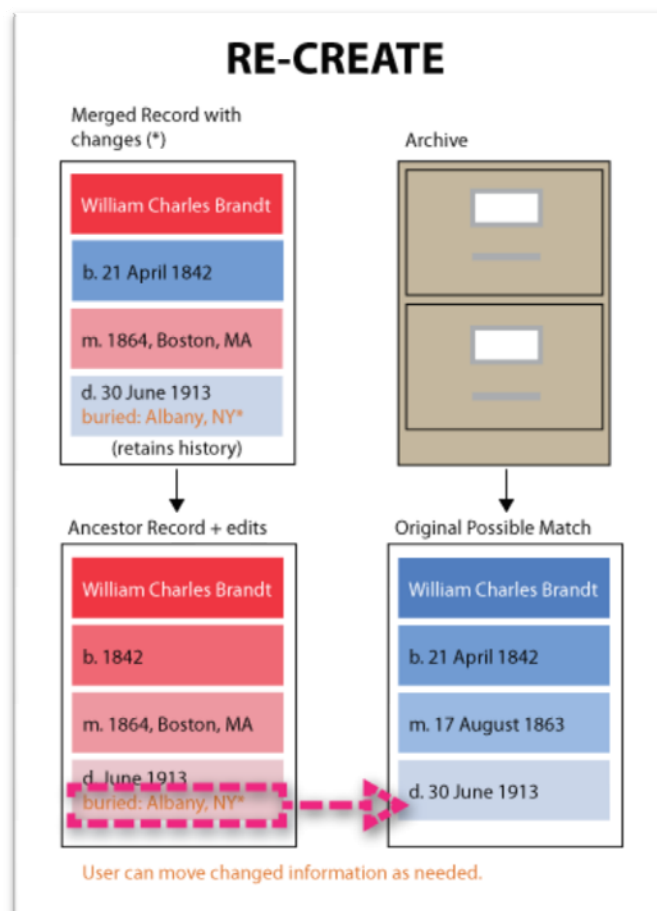Figure 2. How a merged record is fixed if it is changed after the merge.

Re-create the Record That Was Merged and Archived

If someone changes a person after a merge, and if the merge needs to be un-done, then the change history will have an option to re-create the record.

The record needs to be re-created because Family Tree does not know which person the changes were intended for. Therefore, the Family Tree will:

- Retrieve the archived record as it existed before the merge.
- Display the merged record as it currently exists, with all of the changes that resulted from the merge and all of the changes made after the merge.
- Let the user manually move information from one record to the other.

This graphic illustrates how re-creating records happens:

"Not a Match" Feature

In new.familysearch.org, users spend a lot of time separating incorrectly combined records only to have someone else recombine them because the system keeps suggesting them as "possible matches." This can be very frustrating.

To prevent incorrect merges, Family Tree will have a "not-a-match" feature. When someone indicates that two records are not matches,

Family Tree will no longer suggest them as possible matches.

As with other changes, a user who indicates that records are not a match can enter a reason to explain why the records are not about the same person. This will help prevent a lot of bad matches from occurring.

The "not a match" declaration is also recorded in the change history. The change history lets someone undo the "not a match" declaration if it is later discover that the two people really are duplicates.

USING FAMILY TREE TO FIX RECORDS THAT WERE INCORRECTLY COMBINED IN NEW.FAMILYSEARCH.ORG

Because new.familysearch.org lacks a "not a match" feature, users separate records that contain information about distinct people, only to have another user recombine because the system keeps showing the records as possible duplicates.

The combing and separating of records in new.familysearch.org affects Family Tree. After a combine or separate in new.familysearch.org, the Summary tab of a record can change. As a result, in Family Tree, a person's information can be wrong or go missing.

If a bad combine causes a person's information to become inaccurate in Family Tree, users of Family Tree need to correct the situation by:

1. Using Family Tree to correct information about the person whose record now contains wrong information.
2. Adding or finding the record for the inappropriately-combined person, and

making sure that it contains the information it needs.

3. Making a "not a match" declaration between these two records so Family Tree does not present them as duplicates. (*Note:* As of the writing of this white paper, this feature was not yet available.)

CONCLUSION

In new.familysearch.org, duplicates are combined. It is very easy to combine records about different people. There is no way to prevent the wrong records from being combined. Once a wrong combine happens, it is very difficult to fix the problem.

Family Tree will provide a better solution for duplicate records. It:

- Solves the IOUS issue because it does not try to keep all of the duplicate records that are stored within a combined record.
- Lets users correct both inappropriate merges and provides features that allow:
  o Undo merges if no changes were made after the merge.
  o Easily correct and re-create the records when records are changed after a merge.
  o Prevent the merging of wrong records with a new "not a match" feature.

Dealing with Duplicate People