

Data Conversion Project

At the Portland 2011 SFA reunion I heard a plea to convert the SFGS database from Access to a program that could produce a Sheldon Family Book. I got the CD file at that time and started working to understand the data. I called Rose Newton and she allowed me to work on a possible conversion project.

I found the database was in very bad shape. It became apparent that Rose was uncertain what had happened to the data or how it was structured. She also seemed confused on what version of the database I received and if the latest people were included in it. Rose was very good with the paper files and Sheldon numbers but didn't know the folder system well enough to find the most current SFGS database. It became obvious that Rose was a data entry person supporting the creators of the computer database.

The data had been painfully typed into an original DOS program and then converted to other software programs. I was told two other experts in genealogy and computers had tried and failed to convert the data. I realized I was the only person in the SFA who knew Microsoft Access well enough to understand the data format and semantics. I didn't know if the data could be converted or not. I found the file had been corrupted by a power outage which had not been discovered earlier.

Did I say the data was in very bad shape? Let me be more specific. There were over 4000 dates that were impossible. There were words where numbers should go. There were duplicate people in the data. Further, the Sheldon Numbering system was like using Roman Numerals, except worse. There was redundant data that didn't agree with itself. It is apparent that the database creators were aware of those issues as they added computer numbers to try to help some of the problems but only half of the people in the database were assigned these numbers. It became obvious why others had given up.

It became apparent that nobody knew how serious the issues with the data were. I was the only person who knew Microsoft Access well enough to sort and filter the data to find the problems. It was like telling someone they needed oil in their car and had a flat tire, but they kept driving it. Someone had said it was an Award Winning database, so it must have been good, right?

I created a new extended person numbering system and learned an additional programming language within Microsoft Access in my first year. I learned the GEDCOM standard well enough to convert the SFA data into it and import it into RootsMagic (RM). I chose RM because it had a great feature to merge duplicate people. There were over 2000 duplicate people in the data. Many of these had different Sheldon Numbers and/or computer numbers. The SFA board voted to accept my conversion at the 2012 meeting. Unfortunately, the duplicate merge feature in RM wasn't good enough to fix the duplicate people problem.

After the 2012 meeting in Northampton I started again to fix the broken database. This next year found me building tools in Access to insure redundant data agreed with itself. For example, a person's name might be typed in multiple times. I would know that Bob, Robert and Bobby were all the parent of a single child. I did a lot of name cleaning, but always retained alternate names in the data. I needed to add prefix, suffix and nickname fields to the SFA data to accomplish that. I also needed to fix year of birth data, as much of it was missing, when an exact birth date was given. There were at least 20 different cleaning processes I performed using Access programming tools.

It became clearer with each discovery of missing information or bad links that the data was in bad shape. At that point, I suggest that Rose stop entering new data into her SFGS database as I was making so many changes and fixes that I wouldn't be able to do it again, as many of the fixes were manual typing.

At the 2013 SFA meeting in Oberlin, Frank Sheldon and I did some database magic that got us much closer to a new, and better conversion. This time I was able to painfully tag duplicate people in Access before the GEDCOM conversion. I was able to include all Bio Notes for people and import the data into RM. Now in RM we could use its tools to do more cleaning.

In RM we could see over 85 pages of date type problems. RM would tell us a child was born before a mother and other types of problems. This was about 4000 date problems that we could trace back to the SFGS data. It was provably wrong in the SFGS database by looking at the original SFGS data.

I had many phone meetings with Rose to show her how to use RM and how its features were much better than the old SFGS database. I showed Rose how to use Ancestry.com and find fixes to our data. Rose didn't have an Ancestry account and didn't trust its information. She also didn't trust Find-A-Grave.com information and refused to use it either to fix the SFGS data problems. I explained that we knew our data was wrong and we should use any sources available to correct these problems. She felt the only way to find and/or fix the problems was from the original files in her garage. We also discussed how to separate out her many jobs to others so her work was more manageable.

My expectation in converting the data, was to allow others see what had been collected by their ancestors. It had been in Access and was only viewable by Rose and a few other people. By converting it to a standard inexpensive database many others could help discover their roots and help clean or add to the data. I discussed with Rose how a Master database might be used where only a few people would make changes using email sent to then and confirmed for correctness before a change was made.

In December 2013 I downloaded, from Rose's machine, her current SFGS Access database. I used Microsoft Access to compare the original SFGS database with this current version. I found 970 new people in her current version than the older version I had spent 2 years converting. I spent 6 days manually entering these new people into RM and once again told Rose to STOP entering data into the old database. I don't think she understood how hard it was to convert to RM and how I couldn't do it again with her updated file. Did I say how bad the data was?

I have been helping Sue Sheldon with the Database Integrity project, cleaning up data since I turned the RM master over to her. The number of date problems is down to about 35 from the original 85.

I feel that my part of the Database Conversion Project is over. I'm ready and willing to show anyone any and all of my work. Sue, Rose and I have even gone back to the original SFGS database to confirm the mistakes were in the original and not in my conversion.

Board Recommendations:

I believe this data was meticulously collected by H. O. Sheldon and his descendants to help pass along their Sheldon heritage. I believe the creation of a computer database was also to help pass along this information. I think this effort has met a communication roadblock. Rose is the only person who can

look at the SFGS data and I'm the only person in the SFA who can expand on what is in the Access database. (Did I say the data is still in bad shape in Access?)

I recommend that the Board release the RM database to all who might find a Sheldon connection. I don't care if they are a member of the SFA or not. How will others find information if it isn't available. I know this because the paper files and SFGS database haven't been available. How can we see if information is correct, if we can't see what the information is?

I believe that the SFA Genealogist should ONLY determine if an applicant is in the SFA tree. I feel that entering the data and correcting the data and dealing with new applicants is more than a single person can accomplish effectively.

I believe we should move the SFA files to the ACPL so others could see the original information and learn more about their ancestors from original sources. I believe that Rose should do no more database duties and focus her full time effort to help the library move and index the files at their location. Having a library store and care for the SFA files is not only good for others viewing the information but also to guard against possible destruction by accidental fire or other events.

Conclusion:

In the past 3 years I've heard many reasons why we can't show our data to others. I've seen obstacles like needing password protection or login validation before seeing the data. I think it is sad that people told their stories and kept their family records to share with others, all to have them hidden away in a garage and a private personal database. Didn't the builders of the database do it to help share their information? What will the next reason be for not sharing the Sheldon Family history? After working for over 2 years to convert the Access database to a more public structure, why would you not now share your family stories with others?

Marvin Parsons

13 June 2014