



"A German Genealogy Group for Poland and Volhynia"

Submittal Standards of family data for the Master Pedigree Database (MPD)
(Released on March 3, 2011 and replacing Submittal Standards issued on January 9, 2009)

Dear SGGEE Members:

This document is the main document that covers submittal standards to SGGEE. In addition to this document, there are several supplemental documents as noted in the text below.

As the Master Pedigree Database (referred to in this document as MPD) becomes larger and larger, the management of that database, specifically the merging of new data with existing data, becomes more and more challenging.

As more and more members submit data, we see that many of you are making decisions about the formatting of your data that are simply not compatible with easy merging of data. Accordingly, the time has come for us define the problems we see, and also to put in writing how we expect each of you to prepare your data so it is easily used by us who process your data. Before you say that your data has no errors or you give up after reading only a few of the problems we want corrected, be aware that we are also offering methods of solving these problems in THE SOLUTIONS section which is the second bold-face section of this guideline.

THE COMMON PROBLEMS IN MEMBER DATA

1. Location fields that are not true locations Information such as "St. Valentine's Day", "Baptized by Pastor Schmidt", "Died of Influenza", "Home", "70 years, 6 months, 5 days", "Auburn Hospital", "Ballina Nursing Home", "Victoria Gospel Hall", "Rookwood Cemetery" etc. are not locations, yet we frequently see this type of data entered intentionally or unintentionally in location fields. See THE SOLUTIONS section below for a common cause of this problem that you may not even recognize that you are creating.

When a merge is undertaken between the data you submit and the data in the MPD, these locations show up as conflicting data with locations submitted by other members and must be edited out manually for the merge to continue. The same is true when you enter a location of a town or city, but spell the location differently each time you enter it - each different spelling is a conflict. Some conflict in locations is inevitable, given that much of our family data comes from sources of varying reliability, but we really need to reduce the conflicts to the smallest possible number.

Because it takes time to resolve conflicts, we ask that you carefully review the data that you intend to submit to SGGEE and transfer all non-location data to Notes or the proper

location fields of the Events fields and, consolidate the different spellings of location names. Before you start this task, however, please read on to find out what we believe is a fairly simple solution.

2. Given names, surname prefixes and suffixes, and surnames all need to be in the appropriate data fields. When we merge data, we compare surnames with surnames, and given names with given names. If your data incorrectly shows a given name as a surname, then no matches will be found, and we will have spent time doing the merge with no reward for either you or us. See THE SOLUTIONS section below for a common cause of this problem that you may not even recognize that you are creating.
3. Persons with multiple sets of parents We cannot finish the merge process if any person in the database has more than one set of parents. We realize that some people have biological parents and also adoptive parents. If you know both sets of parents, we ask that you record the adoptive parents only in Notes and not link a person to more than one (1) set of parents.

The reason for this essential requirement is that our website has all records linked by families. If a record is uploaded to our website that has multiple parents, our website search routine cannot display any parents of such individuals.

See THE SOLUTIONS section below for our answer to this dilemma.

4. Given names do not match given names provided by others. During a merge of your data with the MPD, the program matches up given names for each corresponding surname. Since it is impossible for the computer software we now use to electronically match up nicknames with full names, or even differences in spelling of the same name- like Adolf and Adolph, Elisabeth and Elizabeth, Ernst and Ernest- we ask that you provide the most common form of a given name according to a list that we have prepared just for the merge of the MPD. See THE SOLUTIONS section below.
5. Surnames do not match surnames provided by others. During a merge of your data with the MPD, the program we use matches surnames using the (American) Russell Soundex System (this is currently the only sounds-like merge available to us).

The Soundex merge compares names like Mueller and Miller (essentially compares spellings without the vowels), but it is generally impossible to compare Polish names to the German version of the same name if they are significantly different - for example the comparison of Czarnecki to Schwarz, Kowalski to Schmidt, Wisniewski to Kirsch or Kirschbaum, etc.

6. The data does not contain dates. If data is submitted to us without dates, then it makes the process of merging data all the more difficult, since there are fewer specifics to compare your data to other data in the Master Pedigree Database. **MORE IMPORTANTLY, if records have no birth date, then we must assume that the person in that record is deceased, and all data for that individual will show on our website. IF YOU DO NOT WANT A SPECIFIC RECORD TO SHOW ON OUR WEBSITE, then you MUST include a birth date that is less than 80 years ago.** We will guard your data, but if you are sensitive about anyone knowing specifics about any individual in your data, then use a bogus date (for birth and if applicable, marriage) to hide that individual on our website. Be aware, however, that if you exclude an individual from viewing by anyone on our website, then

you have eliminated any possibility that other SGGEE members can determine that the person in that record is related to them.

7. The data contains dates and linkages that cannot possibly be correct. Data is sometimes submitted that indicates the birth of a child after the death of the mother, or a birth before the birth of the parents, etc. These errors need to be discovered and corrected before we see the data. See THE SOLUTIONS section below.

THE SOLUTIONS

General comments: Many, but not all of the problems noted above can be traced to genealogy programs that do not follow standard genealogical formatting, and also which use a single data field for the entry of names. We find that Family Tree Maker, in all versions that we have seen to date, is the most frequent problem creator. If you wish to keep using that program, then please read the next paragraph so that you can see for yourself the problems that program generates.

Legacy is the genealogy program that we use to manipulate the MPD data (available free at <http://www.legacyfamilytree.com/Download.asp>). We desire that you also use this program as your only genealogy program, but if you choose to use another program, then we ask that you at least prepare your data using the Legacy program, as it has the capability of detecting many errors that other similar programs are unable to detect- see solutions noted below for specific program capabilities

- a. Legacy has a Potential Problems Report to help you find errors like children born to parents before the parents were born, parents married before age 13, births after the death of the parents, etc.
 - b. Legacy has a Miscellaneous Searches feature to help you discover individuals with multiple parents, to find records of unlinked individuals, and also to find all records with bad dates.
1. Location fields that are not true locations. Legacy has a Master Locations list to help you discover and correct all of the slightly different locations that represent the same location in your data, and also to find those locations that are really notes and events. When you have moved all of the non-locations to notes on each record, we then ask that you correct the remaining locations so that, without exception, you use the three comma system to fully describe where the location is, and in accordance with the location naming guidelines according to http://sggee.org/research/village_guidelines_detailed.pdf You will find additional assistance in finding your locations in the appendix of that document, and also by using our gazetteers at the first item at http://sggee.org/research/village_maps
 2. Given names, surname prefixes and suffixes, and surnames all need to be in the appropriate data fields Legacy has a Master Surnames list to help you discover and correct all of the various spellings you have for the same surname. Seeing a prefix or suffix to a surname in that list (as Family Tree Maker will provide), will also help you find and fix that error.
 3. Persons with multiple sets of parents. Legacy has a feature called Miscellaneous Searches that allows you to find all of the records that have multiple sets of parents. On this same search page, you can also find unlinked individuals, and also records with bad dates.

4. Given names do not match given names provided by others. Given names also need to match up for a merge, and since we do not have the capability to do a Soundex merge of given names, the match needs to be perfect for at least the first 6 letters. Accordingly, we ask that you amend all of the given names for individuals in your database to follow our list of most common German names found on our web site at <http://www.sggee.org/research/AlternateChristianNames.pdf> Please note that unless you change each record, one at a time, you will need a genealogy program that allows global changes in given names to make these changes. Legacy can make such global changes.
5. Surnames do not match surnames provided by others. Once you have made sure that all of your data records only contain surnames in the surname field, you can gain some insight into your surnames, especially if you are using the Polish version of a surname, by reviewing our alternate surnames lists at <http://www.sggee.org/research/AlternateSurnamesDatabaseSubstantiated.pdf>. Remember that if the spelling of a surname does not match anyone else's in our database (or at least a Soundex version of that name), then it is not going to be matched up in a merge of data. The process of looking for alternate surnames should make you a better researcher of your own family by knowing which Polish records may be yours. Please note that unless you change each record, one at a time, you will need a genealogy program that allows global changes in surnames to make these changes. Legacy can make such global changes.
6. The data does not contain dates. All of your records should have dates as best you can determine, but sometimes we realize it is impossible to determine with any accuracy what a date may be. In such cases leave the field blank, but under no circumstances should you enter any words like "unknown" or "deceased" or "not living" or "died in infancy" or "?" in a date field. For living people under 80 years of age whom you wish to remain as not showing on our website, your MUST provide a date less than 80 years ago, even if it is a bogus date.
7. The data contains dates and linkages that cannot possibly be correct. If you use Legacy's Potential Problems report, you can find and correct all of those records that contain data that cannot possibly be correct. Please correct data like children born to a deceased mother, or born before the parents before you submit your data.

While we will make all of the types of changes noted above to as many as a few hundred records in any database that you submit to us, we reserve the right to return a submitted file to the owner for corrections if we consider it to have significant errors that are far too difficult or time consuming for us to correct. We hope that if a file is returned to you, however, that you will make the changes needed and resubmit it to us for inclusion in the MPD.

Thank for your understanding and assistance.

Gary Warner SGGEE Database Manager

Ed Koeppen SGGEE Database Committee Member

Frank Stewner SGGEE Database Committee Member