Place Database

Rob Feist

In family history, almost all information collected answers the questions of who, what, and where. These are the fundamentals of genealogy. As we gather information from sources, we want to be as complete and accurate as possible while maintaining the information as it appeared from the source. Many times the information in a source will need to be read in context to gather the necessary names, dates, and places to continue our research. Having a centralized place database could help genealogists in many ways.

Current genealogical standards record place names as complete as possible from the lowest level place to the highest separated by commas. When viewing places on reports, place names should be spelled out as complete as possible. In a database, however, storing the fact that Salt Lake City is in Salt Lake County in the state of Utah is redundant information. Representing places with their names allows the database to have the same place stored multiple times with only slight variation. For example, St George could be spelled out as Saint George. Other variables include punctuation, capitalization, and even spelling. The place database system must be able to recognize a place in its many different variations in order to reap the benefits from using a central database. Snippets of place names could be extracted from source documents and identified by the system. These places would be checked for validity as a place along with a check to see if that place is a valid choice given the event date. If the place can not be identified, an add request would be generated and reviewed by a data steward for that locale. Once the place is stored in the system, places could be displayed according to user preferences. Finally, a place database would allow an algorithm that merges genealogical data for an individual to use places in addition to names and dates.

An obituary from a local newspaper in Idaho, that references a town named Paris, would be able to be identified by the system as Paris, Bear Lake, Idaho, United States. The context of the newspaper location is critical since there is a city named Paris in twenty seven of the fifty states in the U.S. Place types would not be limited to political places such as city, county, state, province, territory, district, or country. Places such as cemeteries, funeral homes, and religious temples could also be included. Users would be given the option to display these while viewing or printing.

A place database would also bring integrity to the data. A place recorded as Clearfield, Salt Lake County, Utah would be marked as invalid because Clearfield is not in Salt Lake County. Is this supposed to be Clearfield, Davis County, Utah or another town in Salt Lake County? It is impossible to know for sure. A place should also have integrity with the date of the event. For instance, any event prior to Jan 4, 1896 should not be recorded as taking place in the state of Utah since it did not become a state until then. These events should be recorded with a place of Utah Territory.

If a place can not be fully identified, such as Sandy, Salt Lake County, Utah and Sandy is not found in the database, an add request will be generated and reviewed for

accuracy before it is added. Since many databases may be required to handle the volume of requests to identify and lookup places, it is important that there is one central database that is recognized by the genealogical community as the source for place information. Organizations choosing to implement their own system would necessitate translating identification numbers from one system to another. A better approach would have one central database with as many second tier databases as necessary to handle the volume of requests. The databases on the second tier could be a subset of data from the central database. If a database on a lower tier is not able to resolve the request, the request is passed up the chain until it is either resolved or an add request is generated at the central database.

Places are stored in the database as a number, and can be displayed in a variety of formats and languages. Preferences also include the option to display state, province, or country abbreviations. Capitalization and spelling will always be correct since the data stewards check the information thoroughly before adding the place to the database. In addition to having accurate information, place names will also be complete. A city that resides in only one county will always include the county as it retrieves the place name for display. For example, the place Salt Lake City, Utah will always be displayed as Salt Lake City, Salt Lake County, Utah since Salt Lake City is know to be in Salt Lake County.

Lastly, a place database would allow data merging algorithms to use place proximity for an event to uniquely identify a person. Without using places in data merging algorithms, person names and event dates are the only pieces of information that can be used. Including places as part of the identification procedure would greatly improve the accuracy of identifying two records in the database as the same individual.

While there are many different ways in which technology can play a part in family history research, having a solid foundation for representing places is fundamental. It will provide the foundation for automated place identification from source documents, verify accuracy and integrity, and provide the user with display preference options. While implementing this system will not be trivial, support from several genealogical organizations and many volunteers make it possible. This could prove to be a great building block for future technology to simplify the genealogical research process even further.