Open Edit and Source Centric
User Experience Model
for Family History
Collaborative Genealogy Sites
February 29, 2008

Ron Tanner (tannerr@ldschurch.org)

Abstract. This paper offers a proposal for a new user model for collecting, editing and presenting data within a collaborative family history genealogy site. Currently today there are sites that seek to collect accurate data from contributors across the globe in an effort to provide a collective knowledge of information to all users of the Internet. One of the most successful and well-known of these sites is Wikipedia which works to collect and display encyclopedia information from thousands of contributors world-wide. This same wiki model can be applied to genealogical sites working to collect human pedigrees.

1. Introduction

Genealogical sites can benefit from a wiki style user model in order to promote collection of highly collaborative genealogical information from a world-wide audience. Other websites, such as Wikipedia, have been successful in collecting accurate information from anonymous editors, across a broad range of subjects to present a comprehensive store of knowledge to an internet audience. Indeed, today Wikipedia boasts of 2,215,384 articles in its English rendition. In 2006 a review of Wikipedia was done by Library Journal where a panel of librarians evaluated articles on popular culture, current affairs and science. The review concluded that "While there are still reasons to proceed with caution when using a resource that takes pride in limited professional management, many encouraging signs suggest that (at least for now) Wikipedia may be granted the librarian's seal of approval."

Examining a family history genealogical site that adopts a Wikipedia-type user experience model must provide:

- Structure to the interface rather than the free-style format of a traditional wiki system.
- A user experience and rules where registered users can modify existing data and non-registered users can view the data.

- The ability to recover previous values of edited data.
- A model that separates conclusions from evidences in the system.
- Notification and collaboration features.
- Protection of classified or sensitive genealogical information.
- A mediation process to mitigate conflicts between contributors.

Genealogical information on a human-pedigree scale requires the ability to allow many thousands of users to enter the system and easily add or correct inaccurate information. Currently many family history sites are cumbersome in how they allow users to update the data, often discouraging users such that they will create their own entries or give up trying to update the system. What is needed is a structure such that the data and users contributions are safe from deletion and yet the data is easily and directly updated so that users will be drawn to the site and willing to add and correct information. This level of ease and cooperation is necessary to achieve a large-scale, human pedigree site.

2. Structured Wiki Format

Wiki-type sites on the internet today present a free format style that gravitates to a prose presentation of information. Although guidelines are recommended and styles are presented the text is generally loosely compiled and manually linked between articles.

Individual person information in a genealogical system must be analyzable by the system in order to provide validation to assist in accuracy and to generate such presentations as a pedigree or descendancy chart. Additionally, in an effort to reduce duplication, the system must be able to analyze information across individual persons to determine if duplication is present and to recommend merging of individuals.

"Person Pages" in a genealogical system can constitute a collection of data regarding that individual. These person pages can have structured fields where users can enter or modify data regarding that individual and their relationships to other person pages. Having structured fields that result in taggable data allows the system to analyze that data and provide validation such as restricting gender values to male, female, or unknown, and other appropriate rules for the specific field. Since this is a deviation from a strict Wikipedia model, we will refer to this structured model as an Open Edit model.

3. User Modifiable Data

Just because genealogical sites require structure to its data, does not prevent the system from allowing contributors to modify the data with the same freedom that a free-format, wiki-style system provides. Today, many user-reported errors revolve around the inability for the current user to correct inaccuracies on a person or their relationships. Much frustration is felt by contributors when they discover information that they want to enhance or correct and yet are prevented from doing so.

Genealogical data often has more significant emotional attachment to contributors than other types of data, such as the information provided through Wikipedia. There are significant expectations around genealogical data. For example, the National Genealogical Society states in its guidelines regarding the use of technology in genealogical research,

> "…actively oppose the proliferation of error, rumor and fraud by personally verifying or correcting information, or noting it as unverified, before passing it on to others."
> (http://www.ngsgenealogy.org/comstandtech.cfm)

This expectation is also reflected in the code of ethics created for other organizations such as the Board of Certified Genealogists, Association of Professional Genealogists, and other organizations that provide accreditations for genealogists.

With this desire to have very accurate information there will be expectations required for the data placed in a human pedigree site. One could argue that allowing any user to change any data will cause the information to become less reliable. Analysis of Wikipedia has shown that the information tends to move toward accuracy. Despite this, there must be more controls over those that contribute the data.

3.1 Contributor Segmentation

To assist in making the data more accurate, there must be some accountability by contributors on the data that they place into the human pedigree. This can be accomplished by segmenting the system such that only registered users can edit and contribute data.

Registered users must provide some identifying information about themselves and possible contact information for others that wish to

collaborate. Placing one's identity next to contributed information may promote more accurate information, particularly when others within the system can openly question that data and overwrite it if inaccurate. Although this is possible in freely open systems, such as Wikipedia, genealogical sites should require some willingness of contributors to have "skin in the game."

The commitment to provide as accurate data as possible can be enhanced when contributors are known. Consequently, genealogical sites using the Open Edit model should provide customer segmentation – those willing to contribute and face scrutiny and those that are just there to view the information. These types of users can be distinguished by allowing registered users the right to freely modify data while unregistered users can be restricted to a view only mode.

4. Recovery of Edited Data

With any system there is an expectation from contributors that their data is protected such that it will not be lost. Contributors to genealogical sites have a higher expectation of data protection particularly because of the emotional attachments to ancestral data; consequently sites must be particularly sensitive to data recovery. Data recovery is necessary in areas as person page edit, deletions, and merges.

4.1 Recovery of Person Page Edits

Wikipedia and other wiki systems employ a logging system to maintain versions of their pages. Editors of the pages can look at difference logs from previous pages and then restore a page or text from a previous version. With Open Edit, since it uses structured data on the page, recovery of previous field values is considerably easier to present to a contributor than a difference log. For example, the system could show in a table a matrix of data values along with page dates (see sample below). The contributor could restore the page based on a specific date by choosing a column to restore, or set of values by choosing the cell they desire for the field.

Fields	Date A	Date B	Date C
Field 1	Value A	Value B	Value C
Field 2	Value A	Value A	Value B
Field 3	Value A	Value A	Value A

By saving each change to a person page and identifying the specific values of fields, the recovery and restoration of person pages by the contributor should be a straightforward exercise.

4.2 Recovery of Person Page Deletes

There will be occasions when a person page is completely wrong and must be removed from the system. When this is done the system must mark the page as deleted and still retain all of the previous history of that page. The system could also present an interface that shows the pages that have been deleted and allow a patron to restore a deleted page.

When a person page is restored, the system must present the page at it's last saved state and the history of page edits must be visible and usable. Additionally any connections that the person page may have to other pages must be resolved. There could, for example, be a link from a person page to their children, spouse, or parent pages. As the person page is restored, the system must walk the user through restoring these links, making sure that correct relationships are properly restored as linked pages may have been updated with other relationships.

4.3 Recovery of Person Page Merges

As a large system evolves there will always be some form of duplication of person pages that enter into the system. Users may create new person pages representing relatives without searching the system for previously created pages that represent the same person. The system must provide the ability to discover duplicate pages and present them to the user for merging.

When a person page is merged then the user should be presented with an interface that allows them to choose from the two person pages which values that want to have in the resulting combined page. Once completed, the system must display the single combined page. Be aware that relationships and sources must also be merged during this process.

There are often times when a merge occurs that should not have happened. Possibly more sources were discovered that show the two merged individuals were really not the same person. For these cases, the system should provide the ability to un-merge person pages. This could be as simple as deleting the merged page and re-activating the two previous individual pages. There may be a need to identify what values from the merged page go with which separated page as the merged page may have been edited.

There are occasions when a merged page will continue to be merged over time with so many other duplicates that it becomes extremely difficult to properly un-merge the pages. When this occurs, the system could provide a "resurrection" capability that can restore a previously merged person page to its state prior to merging and still keep the merged page in the system. For example, if A were merged with B resulting in C, then C were merged with D to create E, a so forth. Then it is discovered that A really should not have been merged, then A could be reconstituted back into the system without deleting E. This would result in two pages, A and E in the system.

5. Family and Pedigree Presentation

Even though conclusions of an individual person and their relationships would be primarily done on their person pages, the system can still present genealogical data in family pedigrees and group sheets. Navigation through a pedigree or family group sheet has always shown to be a comfortable and effective for patrons.

Family group sheets and pedigrees can still be presented in an Open Edit Source Centric model as a means to navigation and initial entry. When an individual is added to a group sheet or pedigree, this creation would result in a person page with patron conclusions being linked to the pedigree or family group sheet.

Indeed there are many presentations that could be shown to users that present links to person pages and may include relevant family information such as marriage dates and children along with their birthdates, retrieved from person pages. This presents another value for the structured wiki format, as data from the person pages can be retrieved and presented on other pages in a summary type view, such as the family group sheet.

6. Separating Evidences or Sources from Conclusions

Genealogical data is made up of two components: evidences (or sources) and conclusions. Research into historical information often results in several pieces of evidences that provide data about a person or their relationships. For example, birth information may come from a family bible, a birth certificate, governmental or military records that contain dates and places of birth. There is no guarantee that all of these sources will accurately state the exact same information. Indeed, the further back in history we go, the less collaborating evidence we usually find. Consequently, genealogists must

make a conclusion based on all of the currently available evidence. That conclusion must be modifiable should additional evidence be found that warrants a revision of that conclusion.

This paper and others (see Wilson et at., 2006) propose that sources, or evidences, could be separated from conclusions. Indeed, in a non-scientific, brief paper-prototype study it became obvious that contributors view evidence data different than source data, and have a different expectation on how that data can change.

In the paper-prototype study, conclusion information was displayed in a separate window than evidence data. The evidence data included marriage records, death certificates and journals. Several different individuals were interviewed, having a large range of computer and genealogical experience. All individuals were able to recognize the difference between conclusions about a person and the pieces of evidence that support that conclusion. When prompted to make a change, they all were willing to modify the conclusions based on the evidence. When asked about their behavior, they responded that evidence was something that should not be changed but could be enhanced by supplying additional evidences. They did, however, feel comfortable in editing the conclusions.

Separation of evidence from conclusion allows the system to provide different features when dealing with one versus the other. Sources can remain protected and only edited or removed by the contributor where conclusions can be freely edited.

It is interesting to note that those interviewed also were comfortable adding information into the conclusion fields even without evidences, as long as they personally knew the information. However, when the paper-prototype identified that a conclusion value had been changed by someone else, then the response was immediately; who changed the data? and where is their evidence? Those interviewed treated their "living memory" as valid evidence and expected others to show evidence that support or confirms the changes that were made.

7. Driving to a Conclusion

There are many advantages for presenting genealogical information in a system that promotes peer review, source collection, and discussion along with open editing. A significant purpose for a genealogical system is to

collect all available evidence and then work to come to a consensus on conclusions regarding that individual or relationship. Occasionally, emotional and controversial issues arise regarding ambiguous evidence data. Relatives may recall certain events of a deceased relative, documented evidence may be difficult to read or there may be more than one source that state conflicting data.

Genealogical research, although very imprecise at times, has the goal to drive to a conclusion about the person and their relationships. There are often basic rules-of-thumb that have been instituted to help drive to conclusion. In the face of ambiguous data these basic rules could be, for example, taking the earliest birth date as the conclusion date. An Open Edit and Source Centric wiki model promotes this behavior by separating the sources from the conclusions. An Open Edit system can provide "good participation" guidelines, much like Wikipedia does, to promote good citizenship and basic rules-of-thumb to be used within the conclusion human tree. For example, the system could recommend that contributors record, when there is ambiguous data, in the conclusion tree the earliest date of any event. Additionally, they could be instructed that when there is sufficient evidence to conclude a different date than the earliest, then the most supported date should be used.

By separating the conclusions from the source data, the discussions and efforts shift from the conclusion to the evidences. This shift in focus will provide positive reinforcement to research and continue to find corroborating evidence. As additional evidences are discovered, interpreted and added to the system, then patrons will feel confident in their conclusion.

8. Notification and collaboration features

One critical requirement of an Open Edit system is the ability to foster collaboration. The fear of anyone changing data is mitigated if a system can automatically inform interested parties when changes have been submitted. Additionally, collaboration features allow individuals to communicate, discuss and agree on conclusions placed in person pages.

Recently an associate desired to experience editing within Wikipedia. The associate went in and modified a page, not necessarily adding valuable data to the article. Within ten minutes the article had been reverted back to a prior version. Obviously interested parties had been notified of changes to the page, reviewed and rejected the change.

9. Protect Sensitive Genealogical Information

Sites such as Wikipedia are built primarily to share all data with the entire internet community. Genealogical sites generally desire to do the same; however, there exists some data that is considered sensitive or classified. Data that must not be generally shared must be protected either from modification or from viewing. Although most person pages should be fully accessible and open there are some specific cases when a person page may need to be locked. Once such case is discussed in the next section regarding mediation.

10. Mediation

Any time you have an open system that allows contributors to freely modify conclusions, there will be times when two or more contributors disagree on the conclusions that each make, even if they each have supporting evidence for their conclusion. To prevent these person pages from continually being flip-flopped on their values and to promote collaboration and resolution a system of mediation must be in place. Wikipedia provides this mediation capability as well for their encyclopedia pages. This mediation method has been shown to help in resolving conflicts over information disagreements.

An Open Edit system can follow Wikipedia model of mediation and have an escalation path towards resolution. A modified Wikipedia model for genealogical information could be the following:

1. Publish policies and proper etiquette to all contributors that they are expected to communicate with others, provide evidences to substantiate claims and attempt to come to consensus on conclusions. The expectation is that collaborative features within the system will promote this activity. In our paper prototypes when data was changed those interviewed were willing to accept the new data if the evidence supported the information and enhanced the evidence they may have.
2. If consensus cannot be achieved, then collaborators may request an informal mediation. Informal mediation is a process where other contributors that are interested in the same person page can participate in a collaboration process to work through the issues and come to resolution.
3. If informal mediation does not reach a resolution, then a formal mediation may be requested. Formal mediation is provided by volunteers within the system that are willing to participate as mediators in the system. Typically at this point the person page is

frozen (and marked as so) for several days to allow a cooling off period so that emotions may be disengaged in the dispute. A mediator facilitates communication between the collaborators. This can be accomplished by volunteers in the system and may also result in assistance being provided by genealogists willing to assist in finding additional evidences or evaluating provided evidences.

4. If a formal mediation does not reach a resolution, then the matter can be brought to an Arbitration committee. In this committee, arguments are heard and a final decision is made regarding the conclusion.

11. Conclusion

A genealogical site that desires to create a large pedigree must be willing to allow multiple contributors to edit the data within the pedigree. Contributors in the system will generally have the desire to put in proper information and work together to build a successful pedigree. Open Edit and Source Centric models can provide many benefits to the system to ease contributor's interactions with the site by providing the following capabilities:

- Structured data on pages to facilitate system verification and pedigree generation.
- Segmentation of users between non-registered and registered users who are more willing to withstand scrutiny with significant collaboration gains.
- Easy recovery of changed data.
- A model that separates conclusions from evidences in the system and provides different features and protections on that data.
- Notification and collaboration features that promote interaction between contributors along with good citizen guidelines.
- Protection of classified or sensitive genealogical information.
- A mediation process to mitigate conflicts between contributors.

Genealogical sites can adopt wiki-type system models to provide contributors a unique experience, removing barriers for editing, updating and correcting information and enhancing collaboration. Genealogical sites using an Open Edit and Source Centric model can be successful in collecting and supplying accurate information from thousands of contributors working to provide large-scale pedigree information.

12. References

Wilson, Randy, Ouimette, David, Creighton, Tom, Monson, Lynn, Fox, David, Lawyer, Dan, Law, Steven, Heath, John, "The Model with Four Boxes," January 25, 2006, presented in the 2006 Family History and Technology conference.