April 12, 2010

Genealogical Forests and Family Trees

Peter A. and Evan L. Ivie
GeneSys Foundation

Introduction:

Current genealogical programs focus on individuals and the links between individuals. Navigation through these programs involves moving from one generation to another or in jumping to a new individual through a search of some kind. Visibility is provided to the data on each individual and to the parents or children of that individual. These programs have been a great boon to genealogical research because they allow researchers to electronically capture and share genealogical information.

Thesis of this Research:

The question asked in this research effort is whether there are new genealogical tools and techniques that might be developed that can enhance our understanding and ability to visualize the composition, structure, origins, chronology, and geography of a family or group of people. Pedigree and descendancy charts provide some visibility into the composition of a family, particularly if they include date and place information. Are there other techniques that would assist in giving us a birds-eye view of a family? The goal is to make it so that we can see our ***genealogical forests*** and that they are not so well hidden by our ***family trees*** which are so cumbersome to navigate through.

What kinds of information might help us gain a better grasp of our family. Here are some examples of questions that we might want answered:

- How many known relatives do I have?
- How many first cousins do I have?
- From which countries did my ancestors immigrate?
- Where do/did the descendants of a given person reside?
- What was the average or maximum age of my ancestors?
- What was the average marriage age of my ancestors?
- How big were the families of my relatives?
- How many descendants does a particular ancestor have?
- Do I have any relatives that lived (or live) in a given locale?
- How many, and which, of my ancestors need LDS temple work?
- What were the most common given names used in my ancestry?

We do not cover all of these questions in this paper, but we hope that they will bring to mind some largely uncharted territory in the field of genealogy.

Methodology:

We will describe in this paper two techniques designed to help us better see the make-up of a selected family grouping: The ***Relations Table*** and the ***Geo-timeline***. Other techniques are in development but will not be described here.

Relations Table:

Before describing a full ***Relations Table***, let us first describe a ***Direct Relations Table***, a table involving only ancestors and descendants. Each ancestor is grouped in a column labeled by a ***generation number***. Parents appear in column 1, grandparents in column 2, etc. In its most basic form, only one row would exist and the value in each cell would be the sum of the people that fall under that column, as is demonstrated in Table 1.

Generation->	1	2	3	4	5	6	7	8	9	10
Number of Ancestors	2	4	8	16	32	64	112	176	270	376

Table 1. Number of Ancestors by Generation

If you know binary progression you can easily find the first generation of ancestors that is missing a parent. If you don't know binary progression, simply check to see if the number in an ancestor column is double that in the previous column. The difference between the double and the number given will be the number of ancestors not accounted for at each generation. A computer can calculate and show the results as shown in Table 2.

Generation->	1	2	3	4	5	6	7	8	9	10
Number of Ancestors	0	0	0	0	0	0	16	80	242	648

Table 2. Number of unidentified Ancestors by Generation

We can assign additional rows to a particular value or range of values, depending on the amount of screen real estate that is available. For example, if we make people fall into a row based on how many years they lived, we might get something like Table 3.

Generation->	1	2	3	4	5	6	7	8	9	10
Age 0-9									1	
Age 10-19								1	1	2
Age 20-29					1	2	1	2	5	3
Age 30-39						2	4	6	5	9
Age 40-49						4		15	7	18
Age 50-59				2	5	7	11	9	25	28
Age 60-69					3	5	8	32	43	52
Age 70-79		1	3	4	8	10	28	36	51	80
Age 80-89		2	2	6	4	14	25	48	72	97
Age 90-99		1	2	3	5	12	25	15	47	65
Age 100-109			1	1	5	8	10	7	13	20
Age 110-119					1			4		1
Age 120-129								1		1

Table 3. Distribution of Age at Death by Generation

Descendants who lived less than 10 years would be fine, but any ancestors who lived less than 10 years would indicate that some of the data needs more research and correction. In addition, we could easily find the most distant ancestors who lived the longest. We could then compare the same type of table for them and their descendants with another ancestor who didn't live as long to see if that ancestors long life translated into longer lives for their descendants. Using other attributes like birth country or age of marriage for the rows would provide similarly interesting tables.

Most efforts to summarize a large group of people based on lineage focus on either direct ancestors or direct descendants. The next step is to put both in the same chart. Each descendant is grouped in a column labeled with a negative generation number. Children are in column -1, grandchildren in column -2., etc. Column 0 is the anchor or focus point for this chart and represents the individual for which the chart was created.

REMOVALS->	-3	-2	-1	0	1	2	3	
DESCENDANTS	4	1	2	1	2	4	8	ANCESTORS

Table 4. Number of Ancestors and Descendants by Generation

Many people are interested in more than just ancestors and descendants. One example is lineage-linked social networking[1] which is becoming very popular on the internet. The following chart is a concise way to summarize nieces, nephews, siblings and all kinds of cousins in addition to ancestors and descendants. It takes a little explaining, but it is very powerful. The first row is the same as in Table 4.

1621 relatives of Horace Leon Ivie found

REMOVALS->	-3	-2	-1	0	1	2	3	
DESCENDANTS	4	1	2	1	2	4	8	ANCESTORS
NIECES/NEPHEWS	6	15	20	9	23	43		AUNTS/UNCLES
1ST COUSINS	5	55	128	82	277			1ST COUSINS
2ND COUSINS				936				2ND COUSINS

Table 5. Full Relations Table

Next we will focus on columns zero and up for rows 2 and up. They are emphasized in Table 6 to make it easier to find them.

1 Peter Ivie. Lineage Linked Cousins and Social Networking with the FamilySearch API. 2009 Family History Technology Workshop. [Cited: 12 April 2010] http://fht.byu.edu/prev_workshops/workshop09/papers/3-3-PeterIvey.pdf

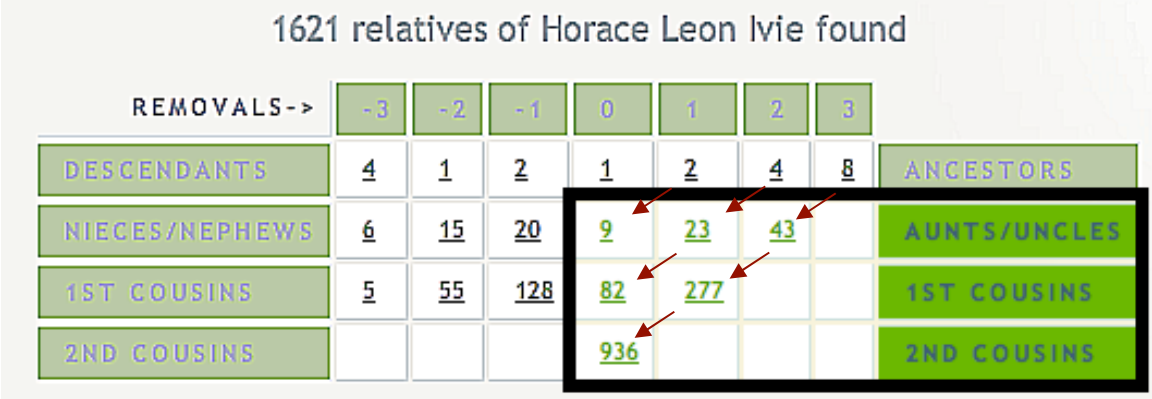


Table 6. Relatives Table with arrows towards children

These cells count the number of children of the people in the cell up and to the right. For example the cell directly below the focus is the number of children of the parents from the cell up and to the right. This particular cell (with a 9 in it) is actually the number of siblings. The cells directly to the right of the 9 are correctly labeled as Aunts/Uncles. For example, under column 2 (with a 43 in it) is the number of Great Aunts and Uncles. The next row shows the first cousins. Column zero indicates first cousins with no removals, column 1 is first cousins once removed, and so on.

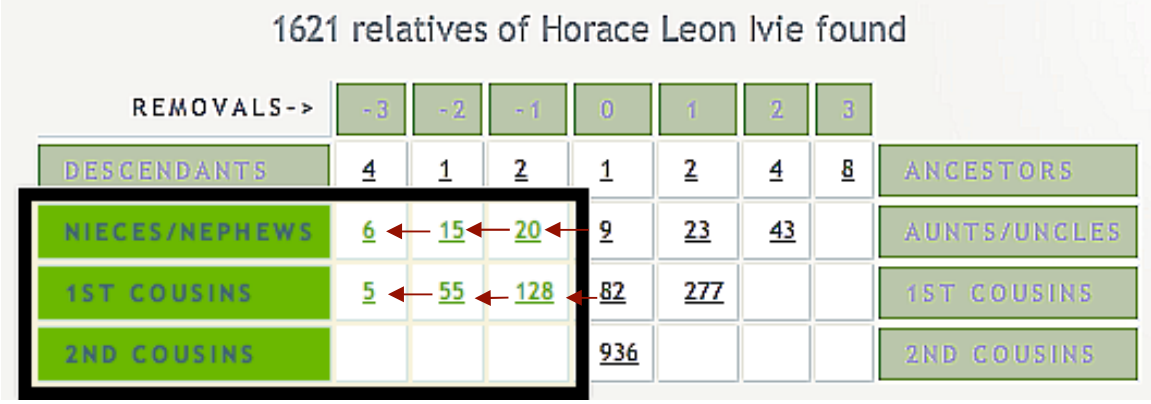
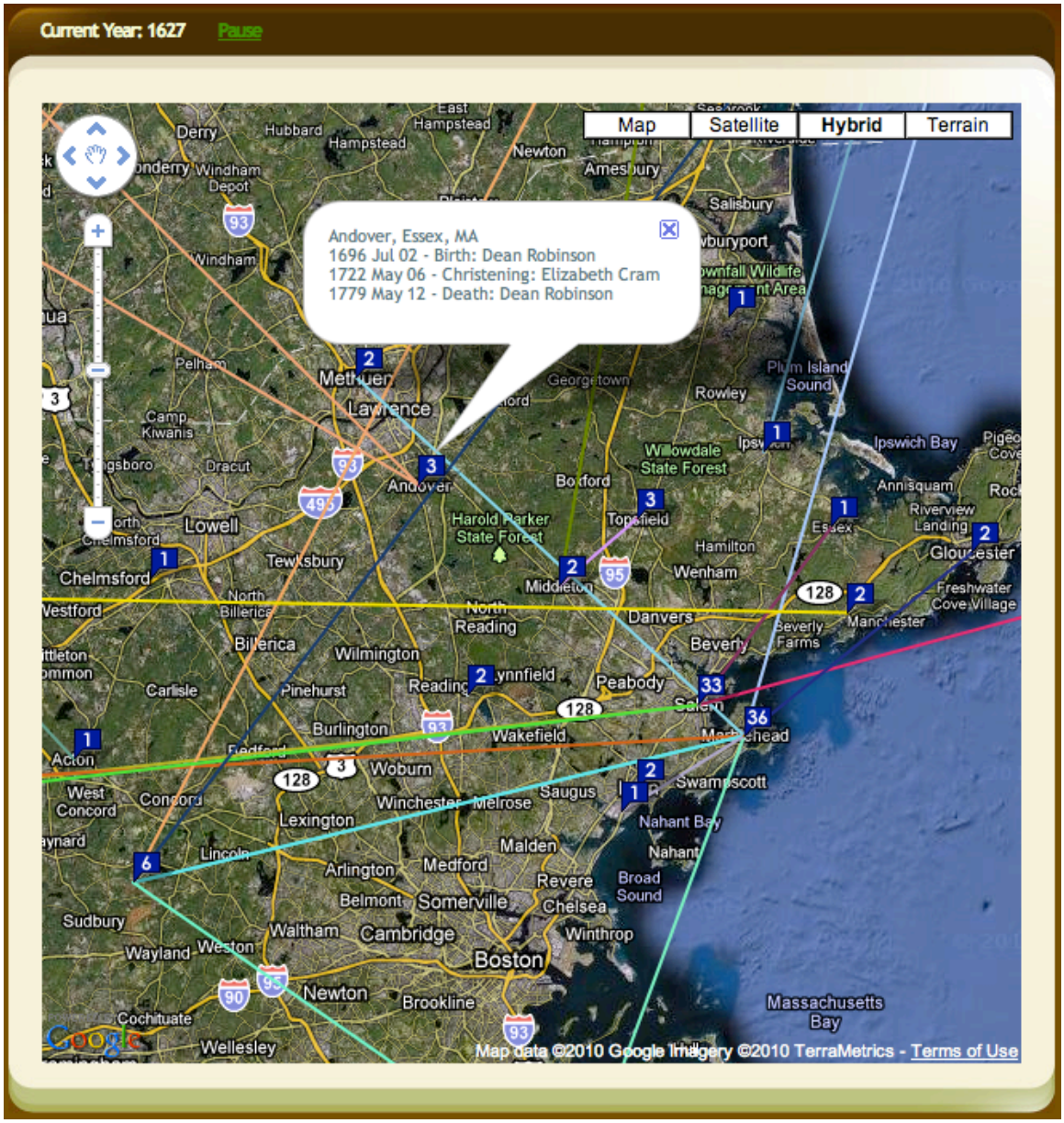


Table 7. Relatives Table with arrows towards children

The negative columns all work the same as the first line. Children are counted in the cell to the left of the cell that their parent is counted in. Originally we wrapped these negative columns into their positive counterparts, putting for example all first cousins once removed together. However, the first cousins once removed who are children of great aunts and uncles (277 of them) are really different than the first cousins once removed who are children of first cousins with no removal. We now call these first cousins with -1 removals (128 in Table 7) to distinguish the two categories.

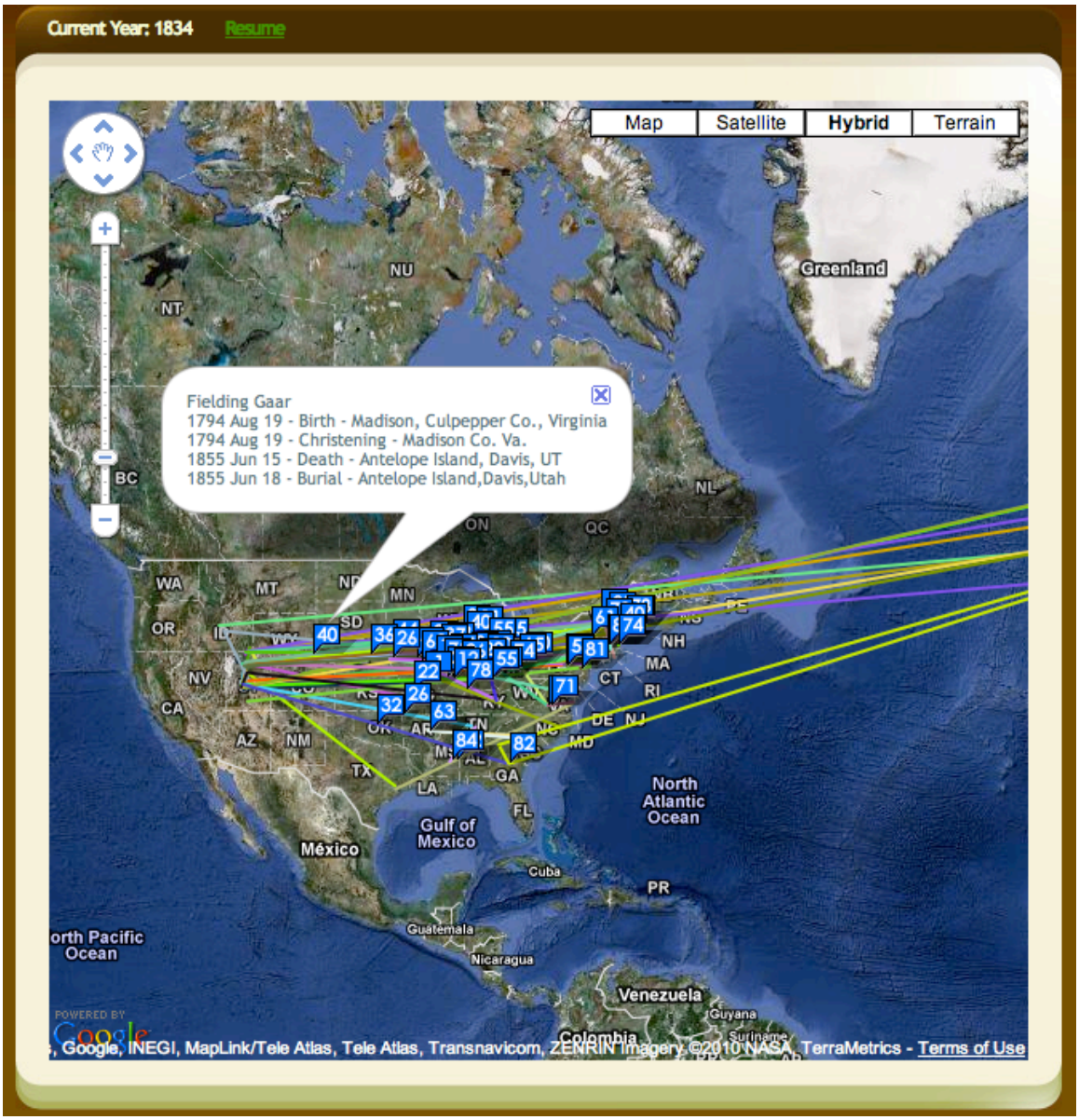
Geo-Timeline:

Another tool that will be demonstrated involves both geography and chronology. Part of this tool is a Google-map which shows where events happened for a given individual or family. The other part of this tool is a timeline for the set of events in question. The user can control a 'clock' which allows time to flow forward or backward. This tool is only partially implemented to date.



Map 1. Showing Events Locations for Relatives of Horace Leon Ivie.

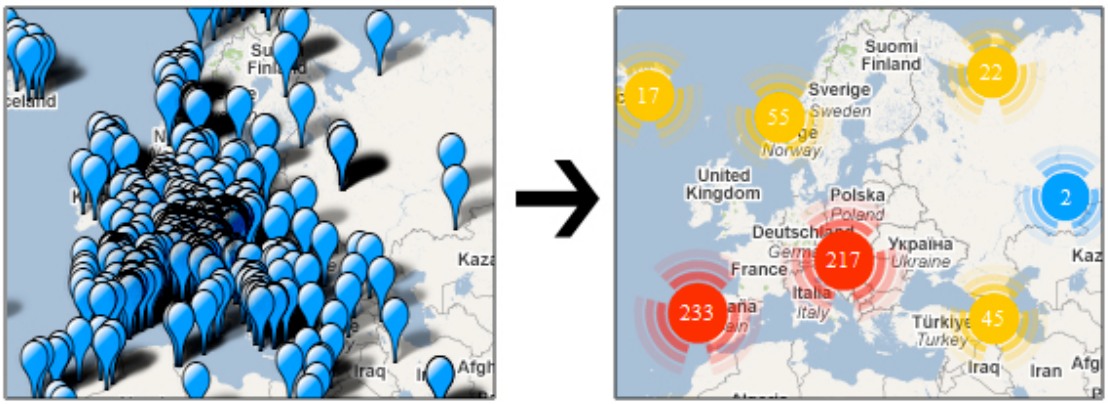
In Map 1, each flag represents a place where an event or events occurred. The number indicates the total number of events over time or only events relating to a person living at a given point in time. It could also be limited to a certain type of event such as births or deaths. Clicking on a flag or drawing a circle around multiple flags could create a list of people connected with the events that occurred at those locations at the relevant period of time. Each person is assigned a color which is used to draw a line between consecutive events for that person. Clicking on the line could trigger emphasis on that path to see all the events for that person. The lines could disappear when the person is no longer living, or they could remain on the map to help get a better feel for the passage of time.



Map 2. Showing Events Locations for Garr Relatives

In Map 2 each flag represents a person living at the given point in time. In this case, the current year is 1834. The number inside each flag shows the age of that person at that point in time. Each flag can be clicked on to show all events for that person. As the year travels back or forward in time, each flag travels along colored lines connecting the events in that person's life. The color of the lines could be random as shown here, or represent the distance between their birth and death, the age they died, or some other criteria. You could draw a box around the lines going across the Atlantic to get a list of the people who migrated to America. The refresh rate could be adjusted to handle high-end computers as well as smart phones. The accuracy could be increased by adding more events to the individuals. The flag could be replaced by an icon to reflect the event closest to the current point in time. For the purposes of a person's approximate location at a given point in time, their events might include the birth of children and maybe even siblings if the person is young enough.

As the number of event markers on a map grows. It becomes increasing difficult to get a good feel for what is actually going on and where the greatest number of events have occurred. This is not a problem unique to genealogy and a lot of research has been done in this area. Perhaps the earliest clustering system was developed by one of the authors while at M.I.T. This system allowed items to be clustered together based on various criteria. A unique thresholding approach allowed control over the size of the clusters.[2]



Map 3. MarkerClusterer[3] clustering of adjacent map markers.

Map 3 is an example of a current product that would likely be particularly effective for genealogy. Markers are grouped together in a "clustered marker" with a number indicating the actual number of markers in the vicinity. Clicking on one of the clustered markers will zoom in on the map and drill down on the markers until you have zoomed in onto individual markers if desired. This could very effectively allow a map to scale for a large number of events. Other solutions to this problem are also available[4].

2 Ivie, Evan Leon. Search procedures based on measures of relatedness between documents. M.I.T 1966 http://hdl.handle.net/1721.1/16395
3 http://googlegeodevelopers.blogspot.com/2009/04/markerclusterer-solution-to-too-many.html
4 http://www.svennerberg.com/2009/01/handling-large-amounts-of-markers-in-google-maps/

Related Efforts:

As mentioned above, pedigree charts and descendancy charts certainly aid in obtaining a more complete understanding of certain aspects of a family. Some pedigrees are color coded so that you can more quickly identify which families crossed the plains, came from Europe, lived in Nauvoo, etc.

Some work has also been done in mapping a given family geographically. The New FamilySearch program has a simple mapping capability. ***Map My Family Tree*** from Progeny Genealogy has some very good geographic functionality[5].

Future Directions:

One of our goals in the development of the Relations Table is to evolve it into a tool for determining relationship to groups of people. Earlier work that we have done involved identifying the relationship of a given person or persons to a group of people [6] [7] [8] [9]. The BYU Roots project was a parallel effort to the work we did [10].

The Relations Table does not have to be restricted to a single person as the focus point. It could have a set of individuals as the focus where the surrounding cells include the sum of all applicable relatives.

It has been suggested that the Relatives Table is particularly well suited to mobile devices. We agree that this would likely be the case, since it allows a large number of people to be summarized and categorized in a small amount of space.

A timeline could be used with or without the map. The most basic form is just a list of events that occurred at the current point in time, with dimmed out events before and after if there is room. Mousing over an event would highlight it on the map if available. Mousing over the name could highlight all events for that person. Another possibility is to have a list of the people living at a given point in time, or a list of those recently leaving or beginning their life based on a given point in time.

There are numerous other extensions to our efforts that we have been working on -- more than we have time or space to discuss.

---

5 Map My Family Tree, Progeny, http://www.progenygenealogy.com/map-my-family-tree.html

6 Pioneer Ancestral Past of the Utah Sesquicentennial, 1997: This program determined relationships about 7,000 request were processed on July 24th and 25th.

7 Nauvoo Pageant Family History Kiosk 1996 to 1999. Provided patrons with their connections to those portrayed in the Pageant.

8 We demonstrated a system that showed relationship to the Jamestown founders at the NGS Conference in 1997 (http://jamestowncousins.com/) and to Notable People of Kansas City in the NGS Conference in 1998.

9 We also offer a system that will tell you your relationship to early pioneers in Nauvoo: http://ldscousins.com/

10 http://roots.cs.byu.edu/

Conclusions:

A version of the Relations Table is currently found at AllMyCousins.com. There are over 3,000 users of that system at this point. We have not yet analyzed and evaluated this user base to determine if they find this tool useful or how hard it is to understand the Relations Table. And the Geo-Timeline tool is new and untested. Are we still in a mode where we cannot see the genealogical forest for the family trees? We are not sure but we are optimistic that these tools can better understand their ancestral origins, descendancy make-up and relatives. Please note that the Relations Table and the Geo-Timeline are copyrighted by, and patents[11] are underway for, the GeneSys Foundation.

11 We have found that it is important for non-profit foundations like the GeneSys Foundation to protect their developments just as for-profit companies do. We are of course anxious that the technologies described herein benefit everyone who would like to use them. However, if commercial ventures incorporate them in their products, we would like to share in the profits that accrue so we can fund further research and development.