The Genealogy Registry/Progeny Link Project

Brief Business Plan

●Quickly Create a High-Quality Genealogy Database of 150 Million Names
●Sell Database Contents for $6 Billion Using Pay-Per-View Methods
●Keep $3 Billion for Investors
●Repeat

Introduction

To keep this document short, I will describe only the most distinctive and striking features of a new genealogical research concept and the computer system that supports it. See the ProgenyLink.com website for online books and other documents that describe the concepts and tools in much more detail.

With new, patented technology, which I can demonstrate and supply, it is now possible to make enormous changes to the business models available for use within the genealogy industry, providing billions of dollars in potential profits where there were none available before.

The traditional methods of doing genealogy research are extremely slow, inefficient, and labor-intensive, and produce mostly low quality results. The long series of small incremental improvements made over the past few decades have been helpful, but have fallen far short of the possible and needed improvements in efficiency. Only a complete top-to-bottom reengineering process can make the extensive overhaul needed to achieve normal industrial levels of efficiency for the current "cottage industry" which is genealogy research.

By markedly changing the methods of research and data entry, and supplying unique Internet-based tools, the entire process of assembling high quality, finished, lineage-linked names in published form for whole nations can now be done 1000 times faster. I define "high-quality genealogy data" to mean data about historical people where most of the usual birth, death, and marriage blanks are filled in, there are links to public records so that the data displayed can be verified, and there are no confusing duplicate entries for that person found within the database.

Industrialization of the genealogy process

The amazing speed and efficiency of the new name assembly processes, and the resulting much lower product price, will make old methods obsolete and will convince most people that they should buy professionally assembled data rather than try to do it themselves. That change in efficiency, price, and consumer attitudes is the source of the new profit potential. This is a little bit like the gradual movement of food preparation from the home to highly efficient food processing plants where finished entrées and entire meals are assembled so that the consumer need merely pop it in the microwave oven for heating. Fast food restaurants provide other examples of out-of-home food preparation. The convenience and low prices of these various food outlets can make home food preparation seem unattractive in many cases, as in homes where both parents work or are simply very busy with other things. The difference in the case of genealogy data is that the price of the new data product can be 1000 times less than the historical cost.

Whether we are building automobiles or picking and processing cotton, the principles of industrialization are the same. Division of labor, specialization, mechanization, and cooperation are the keywords. The same principles can be applied to the assembly of high-quality genealogy data.

As an example, let us assume that we wish to create a high-quality version of the 70 million names of those who died in United States before 1930. Let us further assume that another 80 million names of deceased people will be added to this database from European sources. This 150 million high-quality names should be marketable, on average, for about $40 per name, assuming that most names are sold multiple times for about $3 each. This puts the gross revenue at about $6 billion. If we assume that about one half of that revenue is paid out as expenses, mostly to those who help prepare the data, that leaves $3 billion as net return.

The old methods – Pedigree-sequence research

With old methods of doing genealogy research, the researchers mostly work alone and in isolation, metaphorically wandering through the huge fields of data "cotton," picking a bunch here and a bunch there to add to a node in their pedigree-sequence collection of data. They have no big machines to automatically scoop up whole fields of cotton, "gin" it, and spin out the yarn and weave it. It is all hand-done, very slowly. Since they aren't able to collaborate on any serious scale with other researchers, much of what they do is also duplicated many times over by other researchers, greatly compounding the inefficiencies.

Because of the extreme slowness and difficulty of this kind of work, the final assembled product is often of rather low quality, with only the minimum data included and with only rare references to original source records. A lifetime's work might consist of following every family surname line back five generations. This produces a mere 64 ancestors having 32 surnames. Each generational jump backwards on any particular family surname line may mean a jump to a new set of records in a new location, perhaps even requiring different language skills to read the records. Getting ready for each of these jumps, usually in search of a single person, may require fairly extensive education, and that training requirement alone can make the process extremely slow and inefficient. This whole process is marked by an extreme inability to specialize, since the pedigree-sequence process, by definition, means that no one knows where the next jump back will lead.

The new methods – Descendent-sequence research

With the new descendent-sequence, cooperative research paradigm, the researcher chooses a single surname, presumably his own in most cases, and assembles a descendent structure starting with the most ancient known ancestor. This allows several kinds of specialization which make each individual researcher many times more productive: The researcher need only look for occurrences of one surname in public records rather than attempt to follow potentially dozens of surnames to any spot on the globe. Since families tend to live in a limited area for several generations, the number of record sets which need to be understood and examined is relatively small. This means that the amount of information the researcher needs to learn about localities and record sets is minimized.

The goal of the researcher is to extract every historical person having that single chosen surname and arrange them all into a completed family structure. This extraction and connection process can be 30 times as fast as the pedigree-sequence method of simply looking for one name in a generation. The researcher can take all available related data and interconnect it. It is true that the researcher may not be directly related to most of these people in the strict pedigree sense, but there are many other people, often his cousins, who will have those connections. In other words, if everyone participating in the project is gathering and entering unduplicated data 30 times faster than would happen using pedigree-sequence methods, in the end, the database can be built 30 times faster.

An even larger 1000-to-1 efficiency factor can be achieved as the various individual research groups construct their descendent-sequence surname group data collections. When they have finished their contribution, they can then have access to all the other surname groups which others have built. Each participant needs access to 1024 finished surname groups to complete a full 10-generation pedigree, but only needs to enter one of those groups. All of the needed surname groups should be available by the end of the project. A 10-generation pedigree includes 2048 ancestors having 1024 surnames – which is 32 times the amount of data which most researchers today are able to accumulate and verify in a lifetime. In other words, a 10-generation pedigree would require 32 lifetimes to assemble to high quality standards under current methods, but using new methods it can be done more quickly than the typical five-generation lifetime accomplishment seen today.

Market size

I believe the genealogy industry, and especially its profit potential, is much larger than most people recognize. I estimate that about $66 billion is expended each year in the genealogy industry. I allocate $60 billion to hobbyist or "sweat equity" input, and about $6 billion in actual cash expenditures for equipment, services, training, etc. One important question is how much of this isolated activity can be aggregated, industrialized, and monetized, by replacing volunteer hobbyist activity with much more efficient paid professional or semi-professional activity.

Because of the incredible levels of efficiency which can be attained through industrializing and professionalizing part of the genealogical data assembly process, I believe it should be very easy to turn 1% of that hobbyist "sweat equity" activity each year into paid work for professionals. Even at that minimum level, 1% of $66 billion is $.66 billion each year. Over a ten-year period that becomes $6.6 billion. I believe the number can be quite a bit larger than that, even several times larger, but I will keep the calculations very modest here.

Computations:

Genealogy research labor: There are about 3 billion hours spent on genealogy research in United States each year. (As an example, if 4 million serious genealogists work two hours a day on their hobby they would spend 365X2 =730 hours a year. Altogether they would spend 2,920 million hours. I will round that to 3 billion hours for simplicity.) If we allowed $20 an hour for this intense and complex work, that brings us to $60 billion in effort. Notice that the LDS Church may only have about 1/40 of that workforce, making it desirable to harness the efforts of the entire nation for their projects. We all have the same ancestors, so it makes sense for us to team up to finish the research.

Legacy research duplication levels: Notice that if all 300 million US citizens completed their research back 12 generations using current methods, they would each collect 8192 names, totaling 2.6 trillion names in all. However, there are only 70 million people who died in United States before 1930, meaning that each of those 70 million people would appear in the finished database about 37,000 times. This illustrates the astronomical inefficiency of our current uncoordinated methods, keeping us from completing more than a tiny portion of our nation's genealogy, and shows the need for extensive cooperation and collaboration to make these processes more efficient.

Maximum research cooperation possibilities: If each of 4 million genealogists contributed 18 high-quality names to a cooperative database, they would have completed all of the 70 million names of those who died in United States before 1930. If they spent four hours on each of the 18 names, that would represent two weeks work for each participant.

A huge shift in genealogy economics

As an example of the new economics of genealogy, it should be fairly easy to produce a full five-generation pedigree, consisting of 64 people with 32 surnames, and sell that for $192, assuming $3 is charged for each

name. For a great many serious genealogy researchers, it is considered a lifetime's work to finish five full generations. This new price structure is an astonishing bargain, made possible by the 1000-times increase in productivity.

I don't have all the data I would like on the question of the current effective cost of professional research done in the traditional manner, but I do have a few indicators. I have heard informal quotations of hourly rates from beginning researchers starting at $25. The online sites I quickly examined ranged from $45 an hour to $200 an hour. The minimum starting fee was $50 in one case and $1900 in another. In most cases, there is no way to know in advance how many hours will be required to complete the work. Two websites gave enough information to very roughly extrapolate costs for a full 10 generation pedigree. One would seem to charge $775,526, and another would appear to charge $2,304,000, although I have not attempted to have those estimates confirmed. These two sites had charges by family surname line, but did not mention how many generations they would try to supply. Another site mostly limited their work to 10 or 11 generations, and that may be a common assumption.

One might conclude that the professional researchers would be the group for which the new system would prove most disruptive. The cost for finished segments of the database could easily cost less than 1% of current professional rates. Many of these professional researchers might need to choose whether to use their specialty knowledge to help fill up the new database and collect royalties through that system, or continue to offer their traditional services to a somewhat smaller number of elite clients.

One might predict that a very large number of people in the United States would be willing to pay out the small amount of money needed to acquire a high quality pedigree for themselves from the new database. If 10% of the adults in the nation participated, that might mean that 15 million people would purchase the product. (Some place the number of active US genealogy hobbyists, of various levels of seriousness, at about 12 million, but the number could be larger than that.) At the five generation purchase level that would mean 64 names x$3=$192x15 million purchasers = $2.88 billion. At the 10-generation purchase level that would mean 2048 names x$3=$6144x15 million purchasers =$92.16 billion. There is obviously an extreme range of difference between these two levels of participation. These examples emphasize the potential value of the database, and the need to have some security controls in place to protect the vending process.

Notice that the most extreme value calculation of $92 billion is still only about 150% of the value of a typical year's volunteer genealogy research labor, calculated above at $66 billion, so although it may seem extreme, it is a relatively small proportion over a 10-year or 20-year timeframe. The 20-year volunteer total values of inputs to research activity would be 20x$66 billion=$1.32 trillion, making the $92 billion only 7% of the total.

Cost per name calculations

The cost per name calculation for current professional research appears to be in the range of $50 per name, based on the information from just a few websites. Getting the commercial creation cost down to $20 per name through the new system may not seem like a great improvement, but it is then possible to sell each name (multiple times) for perhaps three dollars each to individuals, which is a 17 times price improvement to those individuals. This should allow people to learn about their roots in an economical way.

But this is only a current snapshot of conditions. Over a long history, the LDS Church has done much to make this relatively efficient research process possible by gathering public records from around the world and spurring other activity in the industry. It should now be time for the LDS Church to greatly improve the economics of its own name assembly processes.

Here is my very rough calculation of the historical cost of processing Church temple names:
1. Records preservation -- assume that an equivalent of $500 million has been spent each year for 100 years. That equals $50 billion. (This involves a guess at the inflation in money values. It also assumes that relative sacrifices made for temple purposes were higher in the past.)
2. Labor on genealogy -- 1.5 billion names have been assembled in a central database. If we assume that each name required an average of 10 hours of research (including completing the related ordinances), that equals 15 billion hours of effort. If we assign a cost of $20 per hour to that work, we get a total of $300 billion.
3. Temples -- about 150 temples have been built so far, times $100 million each, equals $15 billion.
4. Total -- adding all these elements together gives us a total of $365 billion.
5. The Church ended up with 1.5 billion names processed.
6. The cost per processed name equals $365 billion/1.5 billion names = $243 each.
7. However, the cost for each unique name is much higher. With an average duplication rate of 30, the real cost per unique name is 30 times $243 equals $7290.
8. And, going further, the cost for each fully researched and pedigree-linked unique name, that is, those names that are fully connected to other family members, making all necessary temple sealings possible, is probably even higher than the $7290 figure. Perhaps only one third of those unique names have been fully researched and have all family linkages assigned, meaning each of these "perfect" names might have cost $21,870.

The new system I propose should get the marginal cost per name down to the $10-$20 range, allowing the Church to finally receive the full benefit of its immense past investments in genealogy.

Where does the data and money come from?
In building the worldwide database, those who have the necessary genealogy skills and interest can enter all the data to a high quality standard and be rewarded for their efforts. Those who do not have the skills and time to do the research themselves can pay a reasonable fee for the results of that efficient research work. Charging a fee of $3 per name would mean that a full five-generation pedigree of 64 names would cost about $192. Perhaps one half of that amount would go as royalties to the person who constructed and published the data.

The workforce assembling this data can be of many different sizes and types, based on various factors. As one example, if 5000 workers each entered 14,000 names over a three-year period, that would complete the 70 million names of pre-1930 Americans. Those workers could be volunteers, or employees, or they might be considered "authors and publishers," where they prepare the data in advance and receive royalties as their work is marketed through the "pay-per-name" system. As another option, if 100,000 people could be employed – 20 times the staff of the first example – and they each completed 700 names, the work could theoretically be done within 2 months.

Who collects the database value?
We should recognize that if anyone is able to build a high-quality genealogical database which covers the United States and Europe, encompassing perhaps only 150 million names (those being the most valuable names on the planet for genealogical purposes), the resulting product should easily have a commercial value of $6 billion. If the database is later extended worldwide to 1 billion names, and if we use the same calculation methods, that larger database could be worth about $24 billion.

An important question then becomes who will collect the value of that high-quality database? If a commercial company finishes the work, then they are likely to collect the full amount and have a few billion dollars left over to reward their investors. If the LDS Church oversees the project, it appears that they would have the choice of donating that $6+ billion value to the world or to collect that value themselves. In

the gift situation, realistically, most of that $6+ billion value would be collected by professional genealogists who would access the Church data for free and then sell it to their many clients. If the Church were to remove the middleman and collect that money themselves, then they could use the extra money to more than double the size of the finished database rather than let that large sum of money be treated as profits by more nimble companies.

Sunk costs

I can imagine that groups with existing centralized systems will be reluctant to make the radical changes needed to support this new cooperation paradigm. However, the payback for making the change is so overwhelming, that letting past investments prevent the change is not good economics. This is nearly always the case when "disruptive technology" comes along. Those who adopt the new technology can continue on profitably, while those who are bound to the older technology are eventually left behind.

Existing investments and operations

Someone might reasonably ask what effect the changes proposed here could have on existing major suppliers of online access to billions of public records. The expected massive increase in the productivity of individual hobbyists and professionals should greatly lower the cost of the high-quality data product they seek, and also potentially lower traditional subscription income to these companies. However, I believe the raw data which these organizations provide will simply be used much more extensively and thoroughly. Where now a researcher might consider himself lucky to find one corroborating public source for the genealogy conclusions he publishes, it will soon become common to have ALL known public documents linked to the appropriate historical person. That should mean that the vendors of raw public data will have their data used 5 or 10 times as often as at present. This may require them to change their subscription and billing systems in some ways, but their current data should largely retain its value.

Consequences for present genealogy industry groups

1. For the individual genealogy hobbyist or genealogy research professional, there is a possibility of acting in concert to complete high-quality, unduplicated genealogy data up to 1000 times faster than has been true in historical practice. This offers the opportunity to complete the entire United States in less than a year and for each to receive a full 10-generation pedigree. Participants also have the option to earn up to $20, on average, for each high-quality name supplied. For example, by entering 5000 names, the participant might eventually receive $100,000 in royalties.

2. The LDS Church, which plays a big role in the genealogy industry, might have its own reasons to do more than simply focus on the 150 million most valuable historical names and their family relationships. Its interest might extend to preparing a six times larger database of 1 billion unique names from the United States, Europe, and elsewhere, easily supplying 200 years of fully researched and linked names for its temple processes.

With no need to achieve a business profit in its operations, the LDS Church could take the entire $6+ billion in gross revenue from the first set of names gathered and spend it on expanding the database to the maximum extent practical, based on record availability. It could build out the database to marginal cost levels and beyond, where a profit-sensitive company would likely stop far short of that. In the United States and Europe, every historical name properly documented would bring in far more income than it cost to produce. However, it is likely that in many of the Third World countries, the market value of the names would never reach their cost to produce.

The Church could conceivably wait and buy the completed and proven database and then add to it. However, the premium price to the original developers could be rather large. There should be huge cost savings at several levels if the Church manages this project itself.

In the event the Church elects not to do this project, but a private company does, that would still give the Church's members access to high quality names for a reasonable fee from a commercial database to support its temple processes. (Many of those people who purchase names to submit to the temple might also be among those who helped construct the new database, and so are eligible for free access to the names they need, or receive payments for their work which they can use to purchase the desired names.)

Unfortunately, much of the Church's recent online system development work could become underutilized as the private company takes over most of the data aggregation operations. The Church's opportunity to use its volunteer work force to advantage would also be greatly lessened. Its extensive genealogy library system might find itself used even more in conjunction with commercial processes, possibly changing the atmosphere at those libraries.

3. For a commercial genealogy company, it could mean receiving at least $6 billion in gross revenue from sponsoring and facilitating the assembly of this high-quality database. A fairly modest initial investment can be bootstrapped into the $6 billion gross revenue, of which about one half might need to be paid out in expenses, much of that going out as delayed royalties to those serious participants who do the actual work of assembling the data to high quality standards.

Examples of the numbers game for genealogical name assembly

The Huff family in published form and in the public records provides some interesting statistics. The book entitled *Descendents of Engelbert Huff* was assembled in descendent sequence starting with the immigrant ancestor Engelbert Huff, born in 1637, and continuing 13 generations down to today. There are about 15,000 names in the book, of which about 5000 have the Huff surname. This book was assembled through the work of one man, assisted to some extent by several of the Huff cousins, including myself. Its significance here is that in a 10-year period this single researcher accomplished this heroic task. Another remarkably dedicated researcher I spoke with, working about 10 years using the traditional pedigree-sequence methods, managed to assemble the names of about 150 ancestors covering six full generations, plus two lines going back 10 generations. I use these two cases to illustrate the fact that the descendent-sequence researcher was at least 30 times as efficient in assembling names as was the pedigree-sequence researcher. (5,000 descendent-sequence names /150 pedigree-sequence names = 33 efficiency factor.)

Doing large-scale research in descendent sequence has some interesting advantages in today's world where so many public records are available in online indexed form. As an example, there have been about 30,000 people with the Huff surname who have lived in United States, probably with about 6 original immigrant ancestors, and there are about 200,000 public records available which have index entries for the Huff surname. With today's Internet technology, it is easy enough to use a special Internet browser plug-in to examine the online offerings of the LDS Church and Ancestry.com, for example, and automatically locate all of those 200,000 index records, plus the images from which they come, and place all of that information on a PC hard drive. This greatly lowers the individual researcher's clerical burden by collecting in one place all the public information that is available on the Huff surname. A researcher could then either start from scratch to assemble the various sets of descendents from immigrant ancestors, or use this same repository of information to link public records to names which have already been connected in pedigree-linked format.

As a related high-level feature for tracking project progress, it should be possible today to create a "process of elimination" database which has an entry for every known public document and also indicates where in the new database that public document is referenced. In other words, we could know what public data has been used and what has not. This makes the whole process more like a closed system, where there is

clearly only a fixed number of people to document, and we can know at any moment how far we have progressed.

Other computer system features

There are many other patented concepts and specific computer features which are too numerous and complex to cover in any detail in this short paper. There are 27 specific efficiency elements in the computer system, plus several major methodological improvements that are outside the actual computer system.

One important high-level feature is the numbering system which assigns a unique person ID to every person in the database. The multi-part number identifies the supplier of the genealogical data, and also the surname group to which that name belongs. This maintains the personal responsibility and recognition for those who enter and improve the data, and also allows entire descendent-structure surname groups to be treated as separate logical units of data where that is useful.

This numbering system also allows any number of websites on the Internet to be treated together as a single integrated genealogy database, with text and multi-media data concerning each unique person residing wherever it is most convenient. Dispersing voluminous data over multiple web sites can greatly reduce the required size of the central server. Short Internet links to external data could replace voluminous multimedia storage at the central site.

As a curiosity, we might note that many of the efficiency gains of this new system could have been accomplished using a manual system with new procedures 100 years ago if someone had thought of it. The "computer" would have consisted of a room containing 2000 4-drawer file cabinets plus 500 clerical workers. Completing the basic United States genealogy would have taken 17 years to accomplish then, while today's computers could accomplish it in a few weeks.

Summary

This paper has emphasized the practical efficiency, and therefore the cost efficiency, of reengineering one important aspect of the genealogy industry, the actual process of assembling finished, high-quality historical names into an integrated database containing no duplications. The concepts of division of labor, specialization, and "industrial strength" cooperation, made famous by Henry Ford and others, can finally be applied to the genealogy data assembly process, bringing efficiency improvements of 1000 times over past practices, as is typical in most industrialization projects. This should help individual researchers, hobbyists, and those with only a casual interest in their genealogy, as well as institutions such as the LDS Church, to all quickly and thoroughly accomplish their goals concerning collecting genealogy data. This might be accomplished through purely commercial means, with generous rewards to the investors, or it might be done more extensively through the involvement of the LDS Church.