20120914 its all about the numbers for genealogyV04-trim

The Genealogy Registry/Progeny Link.com Project

As with many business deals

It's all about the numbers

although this is a little more complicated than corn futures

At the end of all these computations there is

An easy harvest of $3 billion:

with more processing and income available if desired.

The Basic Records

1. The Basic Records – numbers worldwide

The LDS Church, which takes these matters very seriously, has estimated that from 70 to 90 billion people have lived on the earth at some time. For about 10 million of these people, there are existing documents naming them in some way, at least once. In fact, on average, these people have been listed about six times, meaning that there are about 60 billion record entries for them somewhere in the world. With an average of about two entries about people per physical document, there are about 30 billion record images which exist somewhere on the Earth today.

The LDS Church has plans to eventually collect all available record images. Since it began work in 1938, 74 years ago, it (plus other organizations) has collected about 20% of the world's genealogically significant documents, giving us about 6 billion images containing perhaps 12 billion entries concerning people. In other words, using the LDS Church's worldwide estimates, the records processed so far should give us data on 2 billion unique people. However, I believe that in the Western world people have at least twice as many record entries per person as those in the rest of the world, so I am going to estimate that we actually have captured (imaged) and centralized data for only 1 billion unique people.

The LDS Church is imaging about 100 million new documents each year, indicating that, at current rates, it will take another 240 years to finish that labor-intensive task for the remaining 24 billion documents. With today's technology, this process could be speeded up through individuals using personal and mobile devices to capture more original documents in more places, although this step is not necessary for the process I propose.

2. The Basic Records -- record processing and availability

It is all very well to have some of the world's most valuable documents copied and centralized, but how well can we use them? The LDS Church estimates that 23% of its 3.1 billion microfilm images have been digitized, and that at current rates it will take about another four years to finish that digitizing task. Of those 3.1 billion images, about 600 million have produced about 3 billion searchable names. The

transcription process is completing about 300 million images or 600 million names per year. This should mean that the transcribing process could be done in about 8 years.

(I am emphasizing Church statistics on the handling of records, partly because the Church makes the statistics available, where other organizations may not. In other words, the LDS Church becomes a proxy for the activities of much of the rest of the genealogy industry.)

Ancestry.com and other commercial organizations also have large masses of data which have been imaged, digitized, and indexed and placed online. If Ancestry.com has about 8 billion indexed records for historical people online, those 8 billion records probably mostly cover the same 1 billion unique names estimated above. There is probably much overlap between the records of the LDS Church and of Ancestry.com and other such commercial organizations, but I haven't attempted to estimate those overlap rates.

With 3 billion searchable names from LDS Church sources, the rest to be done in 8 years, and about 8 billion indexed names from other organizations (with much collection overlap), now is a good time from a business standpoint to gear up to completely harvest the data for the 1 billion people described in the records that have been collected so far by multiple organizations.

3. The Basic Records -- the costs of record processing and availability

A. Although it's very difficult to estimate, I'm going to guess that the cost of preparing the genealogically significant records which remain on the Earth is about $250 billion over the past 400 years. Of course, no one alive today paid the cost of creating those records, so they are mostly free if they still exist.

B. Making copies of those records is definitely not free, and I will guess that the Church has spent about $0.5 billion a year for the last 50 years, giving a cost of about $25 billion. The Church makes these records available for free to the world.

C. Getting those records into a form usable by computers is also not free. Digitizing, transcribing, and indexing those records has probably cost someone, notably the LDS Church, about $5 billion.

Assembling All Those Records Into Families

4. Past and current Church efficiencies and costs for names and families

The LDS Church has historically had low efficiency and high costs for the names submitted for temple work. By one series of estimates of costs, including labor costs, each name submitted to the temple cost about $243. But because the duplication rate is about 30 times for each name, the real cost for a unique name is about $7,290. Because only a small portion of the names submitted are also connected into families, which is necessary for completing all necessary temple work, the real cost for each perfectly completed name is much, much higher. I believe that far less than a third of those names (perhaps as few as 5%) have been connected into families, so the real cost for each completed name is at least $21,870. A reasonable goal using today's technology would be to put the cost of perfectly completed, family-connected names into the $10-$20 range.

The LDS Church has been spending perhaps $100 million a year on computer software and database development for more than 10 years, but somehow it has never seriously addressed the problem of reducing the cost of correctly and completely assembling names into families. It has focused the bulk of its efforts on simply cleaning up the chaos in its past temple data.

(We should note that the LDS Church seems to feel that it has little practical incentive to improve upon its past performance in assembling families together. The reality of the temple ordinance process is that the patron time required for each individual historical person is about 90 to 120 min., while the time required to seal an individual into a family is only about 1 or 2 min. Since the overwhelming requirement is to fill those two-hour time slots, and that is the overwhelming bulk of what brings people to the temple, then the small, separate, minimal time requirements of sealing families can be nearly dispensed with. This is a serious departure from formal doctrine, but the practicalities of filling the major time slots seem to be ruling at the moment.)

The Church has recently found ways to lower its costs for names to feed the temples. It's Online Indexing program has given it an infinite supply of lists of names to send to the temples. Apparently that same system also supplies large numbers of names for the sealing ordinances. There are many cases when spouses can be sealed together and children can be sealed to their parents, and, occasionally, grandparents might be included. This is a reasonably good stopgap measure for supplying temples of names. However, there are still many things missing from a complete and correct treatment of historical names. In the area of sealings, the census forms can only supply part of the data needed for one, two, or maybe three generations of family links. And, of course, the census records only show those who are currently living at a particular address. Older children who have moved away will not be listed with that family. Any children that were born and died between censuses will not be included. In other words, census records are typically not complete. It is helpful to seal together nuclear families, but there is no process going on whereby longer strings of family connections can be established, for example five or 10 generations connected together. And then there is the problem of repeating people many times over as they appear in multiple censuses. If one person appears in six census records, then that person will probably appear as six different people in the Church records. It is only by using a more thorough process that these duplications and omissions can be avoided.

In other words, the completeness and accuracy of today's Church genealogical database is only slightly improved over the historical experience described above. It is probably rare to find families complete at each generation, and with numerous generations all properly tied together. This probably means that the costs for each properly completed name are only slightly lower than the historical costs mentioned above.

If the LDS Church adopted the methods which I am proposing, its costs could be much lower and the benefits to it much higher than if this process were implemented commercially. However, there appears to be no serious chance of the Church adopting these methods anytime soon, leaving a clear path open for a commercial solution. The two patents which have been applied for, with one already granted, should offer significant protection against the Church technology people trying to adopt this intellectual property without suitable compensation.

5. Vast new efficiencies with new procedures and tools

The traditional methods for assembling separately listed and identified individuals into families has been extremely slow, cumbersome, and inefficient. Researchers typically have to work almost completely alone and have little or no opportunity to coordinate their work with others. This inability to cooperate in any effective way is demonstrated by the 30 to 10,000 duplication rates for unique names found in historical Church genealogical databases. Duplication rates would reach the astounding level of 37,000 times for each historical person in the theoretical case of every US citizen doing a 12 generation

pedigree using traditional methods. In the past, there was no good solution to this extreme inefficiency. However, today we have a whole series of easily-applied industrial-style cooperation efficiencies which can reduce duplication rates to zero and then increase positive productivity by up to 1054 times.

To begin with, these giant leaps forward in productivity are only available to a trained group of professionals using specialized tools. Eventually, these methods and tools could be made available much more widely in the industry so that the extremely inefficient traditional methods disappeared for most purposes. It is through applying these new principles that costs can be lowered and convenience increased so that the operators of the industrial process can collect large rewards in the range of $3 billion.

5. Market – $66 billion annually, $660 billion over 10 years

Actually, this $3 billion I describe is a very tiny portion of the entire genealogy industry. I put the financial measure of the genealogy industry at $60 billion a year for hobbyists' labor and about $6 billion year for actual cash expenses for training, conferences, travel, data subscriptions, computers, Internet connections, etc., totaling $66 billion each year Over the ten-year life of a project such as I suggest, that becomes $660 billion in economic activity. I am suggesting that we only need to transform 0.5% of that enormous amount of activity into company income by replacing hobbyist activity by professional activity which provides a great cost savings and convenience. The opportunity is there for that 0.5% to become 5%, a 10 times larger number, making $30 billion over 10 years still be a very tiny slice of that market size.

I mentioned that there are probably records available now which would support assembling the names of 1 billion historical people into families. If these connected names are marketed for $3 apiece, and are sold an average of about 14 times because of the way genealogies are constructed, that gives us about $40 per name as gross revenue.

But notice that I am only suggesting that we do a tiny part of that 1 billion names, the most highly valued names consisting of the 70 million people who died in United States before 1930. Assembling a high-quality database of those 70 million names is the course I suggest for reaching that $3 billion in gross income, with up to a 40-to-1 return on investment. (The figure of 70 million names might be bumped up to 140 million names to include 70 million names from Europe.)

It then becomes optional whether to continue to assemble the remaining 930 million names that are in accessible form at the moment, or to go on even further and finish the 2 billion names which could become available during the 10 year life of this project I propose.

It is important to note that the LDS genealogists probably comprise no more than 2% to 5% of the nation's genealogists. They are an important market because of their enthusiasm and determination, but they are still only a tiny fraction of the whole.

6. Result: $3 billion in gross income, or about $2.9 billion in net income.

Implementation Options

Large
The "large" version focuses first on the new procedures, but also requires a full version of the Internet tool. It makes sense to begin the discussion of implementation options by considering the most aggressive option, requiring the most capital. It is the simplest and most straightforward method which allows all of the efficiency tools and procedures to be quickly implemented, the database to be quickly finished, and the profits quickly collected.

After the pilot project is running satisfactorily, it would then be possible to scale up quickly to full-size. That could mean spending about $70 million on creating the full database. If all parameters are confirmed as the project progresses, that would bring in $3 billion in gross income, leaving $2.93 billion in profits.

Small
The "small" version focuses on the Internet tool. The hope is that the features which are offered to the public are sufficient to get numerous people using the system. At some point they may come to understand the many powerful features of the software.

The downside of starting small is that it does not include a large demonstration segment of the project which is designed to convince the entire genealogy industry that there is indeed a strikingly better way to do their research work. Without demonstrating the 1000-to-one productivity improvement which is possible by using the correct procedure AND the correct tool, the users may never come to understand the real intent of the system and never take advantage of it.

Medium
The "medium" version of an implementation strategy would start with a large demonstration project, spending at least $200,000 and perhaps $1 million to build a very sizable database to make clear to the genealogy industry that something quite amazing is now possible. (A first increment could be as small as $20,000 to demonstrate internally the power of the system.) At that time, the hope would be that large numbers of people in the genealogy industry would choose to join in the project and publish their high-quality data on consignment to the central publishing project. Eventually these participants would have to advance a large amount of time and work, possibly reaching a value of $70 million, with the faith that they would eventually receive very adequate compensation in royalties for their high-quality work.

The World Genealogical Record Situation

We have data on 1 billion unique individuals captured and in suitable form for computer processing.

However, none of those names are in a high quality database, family connected, without duplicates, etc.

Now is a good time to set up the system to harvest this data profitably.