The LDS View of Today's Genealogy Computer Problems

Solving The "Sealings Gap" at Temples

It appears that only about 5% of names submitted to the temples are family connected, meaning about 95% of those names are NOT family connected and sealings cannot be done. Today's methods do not supply a complete solution to the scriptural requirement of "a book ... worthy of all acceptance." D&C 128:24.

The "Sealings Gap" Problem

Image: The top floors of the temple, the sealing rooms, are missing in this picture, and that space is empty to symbolize the fact that perhaps 95% of the names processed through the temples are individuals who are *not* connected with their other family members -- their parents, spouses, and children, and their names never reach the top levels of the temple where sealings can be done. For example, the 140 million names in the 1940 U.S. Census are all entered into Church computers to be processed through the temples as individuals, and essentially none of them will have their family connections, meaning their final family-linking ordinances cannot be performed.

The new methods and computer system proposed here will allow the necessary family connections to be made up to 1,000 times faster than it is done today, so that the "sealings gap" can quickly be closed. This new mechanism will also avoid all the duplication of names being submitted to the temple. For example, in the case of the 1940 Census, at least 80% of the 140 million names will be duplicated multiple times in the Church ordinance system files unless appropriate research work eliminates those duplicates. Since people's names may appear about six times in census records and perhaps six times in other public records, the only way to prevent layer after layer of duplication in temple processing is to first tie all of these public records to the single appropriate unique name. All of this needs to be done correctly BEFORE submitting names to the temples, not afterward.

A Broader View of
Today's Genealogy Computer Problems

The World's Records and People Situation

"70-90 billion people have ever lived. 10 billion have been documented and 60 billion records exist for them. "

(Based on broad estimates from different sources, chiefly http://ancestryinsider.blogspot.com/2012/08/rod-degiulio-all-about-records.html)

World record entries
--60 billion

This graphic is meant to show that only a microscopic portion of the available world records have been put in finished form, that is, unique individuals (no duplicates), all reliably identified and placed into family-linked form.

Here Is The Biggest Problem!
And there is a complete answer!!

World record images (1 image for 2 people, on average)
--30 billion

Unique historical people-- 10 billion documented somewhere in the world
(6 record entries for each person, on average)

The "finished" high-quality, unique name count is near zero and is unacceptable:

Only about 0.05 billion (50 million) unique individuals are found in multiple, low quality, highly duplicated databases -- partially family-linked.

11 billion Captured

6 billion Captured

10 billion individuals appear in the world in documents

1 billion captured

As the associated graph illustrates, of the 70-90 billion people who have ever lived on the Earth, about 10 billion of those people are described in a physical document somewhere on the Earth, and, in fact, on average, there are six documents for each person, meaning that about 60 billion existing records need to be captured and processed to be sure we have covered all possible people for which there are records.

The graph is meant to illustrate how far the various record processing steps have progressed. It appears that through the efforts of the LDS Church, Ancestry.com, and other organizations, we have captured about 6 billion of the 30 billion record images in the world, which means we have captured about 11 billion of the total 60 billion name entries that exist, which should now give us centralized access to about 1 billion of the 10 billion unique individuals who are documented somewhere. After spending the equivalent of perhaps $100 billion by the LDS Church since 1938, we have completed up to 20% of some of the basic processes.

Visible versus invisible
The various processes of finding the world's records -- getting permission to make copies, actually making the copies, putting them in digital form, and indexing them -- are all difficult, time-consuming, and expensive. However, they do have the convenient characteristic of being visible and measurable and relatively straightforward.

Managers and workers who have no special knowledge of mathematics or computer systems can function perfectly well in these beginning steps of data capture, digitizing, and indexing. However, when we get to the last step of finishing all this work and placing it in a high-quality database, there is no substitute for a clear understanding of computer capabilities, genealogy mathematics, and also principles of high-efficiency mass production methods for cooperative assembling of data. I assume that this lack of top management's special knowledge of mathematics and computer systems is a major part of what is keeping this final step from being properly solved.

The main focus of this graphic is to point out the microscopic number of names which have proceeded through all the steps until they are in finished form. That is, they are in a database, accurately identified, without duplicates, and connected with all the appropriate family members. The early steps of finding records, copying them, digitizing them, indexing them, etc., have proven to be relatively easy in comparison to the problem of the final selection and assembly of names into the desired finished, unduplicated form.

Just to point out the obvious, if we cannot put all these captured and capturable names into finished form, then a very large part of the preceding straightforward industrial processes of gathering and indexing those names will be for naught.

This whole long process seems to break down at the end, where machines and simple clerical tasks cannot be relied upon to produce an acceptable product. Of the approximately 1 billion names for which we have at least one document available, it appears that there are only about 0.05 billion (50 million) names that have been put into a form that begins to approximate the desired finished database. And the real number that have been put into final, high quality, unduplicated form in a central database is closer to zero. Some of the shortcomings are: Hardly any names that appear on public databases have accompanying documentation that allow quick verification, the LDS Church FamilySearch database is notorious for its high levels of duplication and confusion, etc.

If we cannot get this final step working quickly, efficiently, and accurately, then the vast amounts of time and money spent on the earlier processes of raw record acquisition will never bear the intended fruit. One might even conclude, from watching Church processes, that it has been determined that this final step can never be solved, so all efforts are still being focused on the early stages.

It is remarkable how much effort volunteers are willing to put into the early processes, especially including Online Indexing of these images, even though there is currently no plausible expectation of creating a high-quality central database as a result. Individuals can continue to piece together their personal pedigrees, but there is still no acceptable central place to store and share their work. They may have high-quality data on their personal computers, but that quality is severely degraded as soon as it is placed in any of the current central databases.
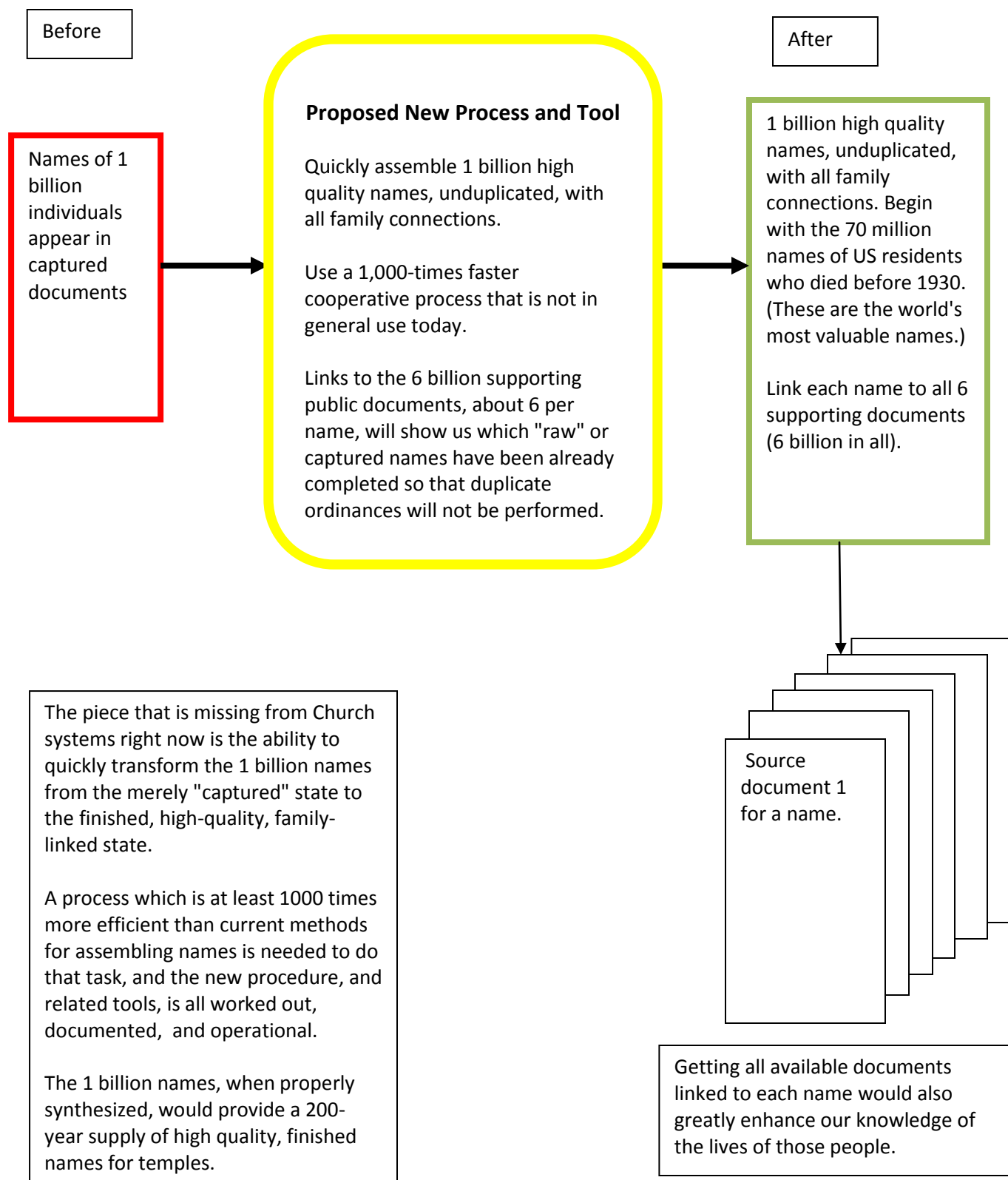
If we cannot get this "finishing" process working quickly and reliably, it seems possible that the enthusiasm of the volunteers will gradually wane, and at some point, even the relatively simple Online Indexing processes will cease to be so enthusiastically supported.

The institutional Church seems to be willing to limit its expectations to simply having enough names to keep the temples busy, even if at least 80% of those names are duplicated five or six times, as multiple sets of records are submitted to the temples in raw list form without first connecting them into families. If all public records for a person were linked to that person, it would then be feasible to completely eliminate multiple submissions for each name.

What Processes and Tools are missing?
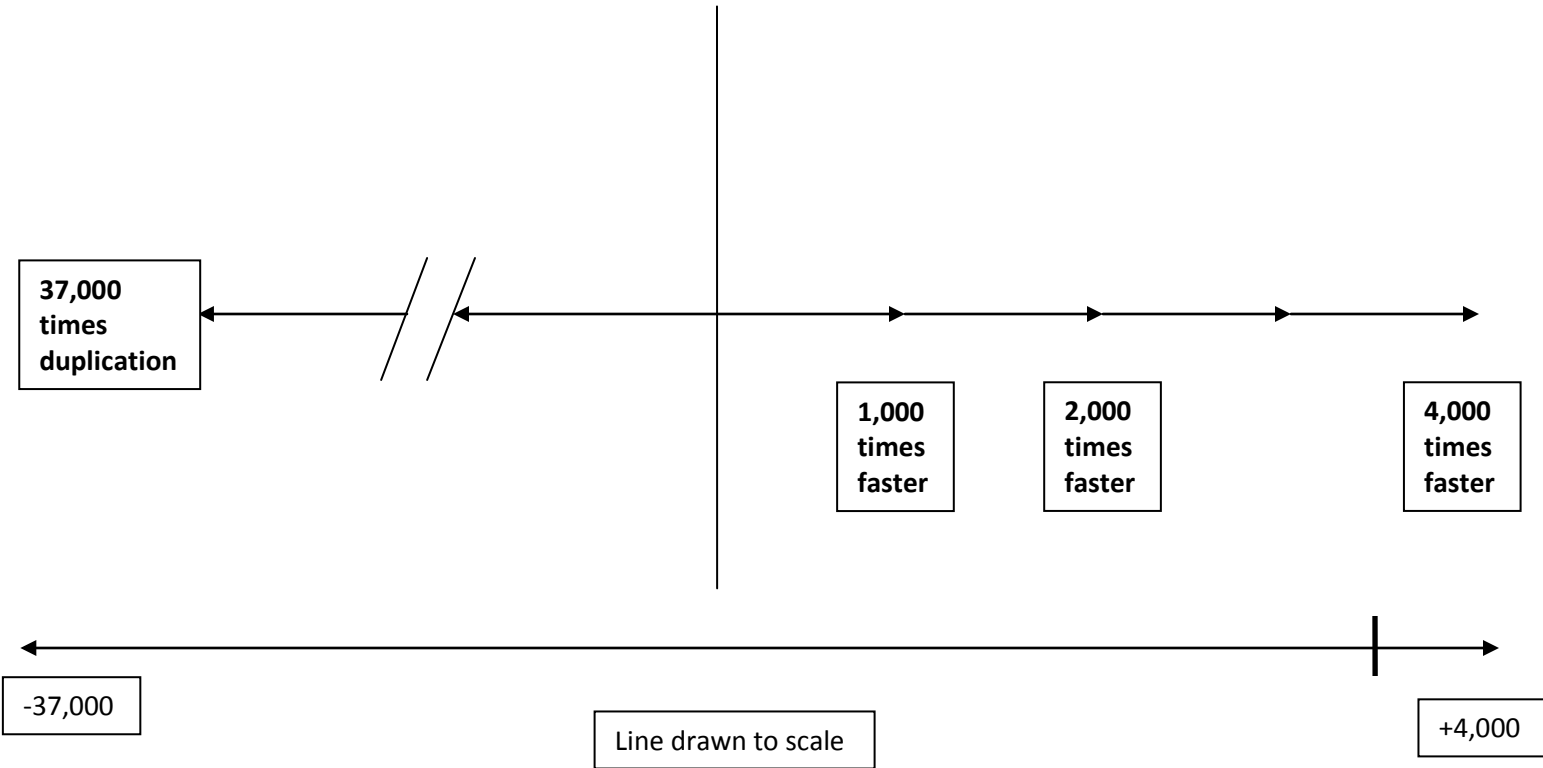# The Answer:

Genealogy Registry/ProgenyLink

Before

After

Names of 1 billion individuals appear in captured documents

Proposed New Process and Tool

Quickly assemble 1 billion high quality names, unduplicated, with all family connections.

Use a 1,000-times faster cooperative process that is not in general use today.

Links to the 6 billion supporting public documents, about 6 per name, will show us which "raw" or captured names have been already completed so that duplicate ordinances will not be performed.

1 billion high quality names, unduplicated, with all family connections. Begin with the 70 million names of US residents who died before 1930. (These are the world's most valuable names.)

Link each name to all 6 supporting documents (6 billion in all).

The piece that is missing from Church systems right now is the ability to quickly transform the 1 billion names from the merely "captured" state to the finished, high-quality, family-linked state.

A process which is at least 1000 times more efficient than current methods for assembling names is needed to do that task, and the new procedure, and related tools, is all worked out, documented, and operational.

The 1 billion names, when properly synthesized, would provide a 200-year supply of high quality, finished names for temples.

Source document 1 for a name.

Getting all available documents linked to each name would also greatly enhance our knowledge of the lives of those people.

Improving Efficiencies in The Genealogy Industry
By At Least 1000 times
Using Industrial Mass-Production Techniques

Remove Negatives: Massive Duplication Inherent in Current Research Methods

Add Positives: Increase Efficiency and Productivity up to 1,000, 2,000, or 4,000 Times Through Multiple Techniques

37,000 times duplication

1,000 times faster

2,000 times faster

4,000 times faster

-37,000

Line drawn to scale

+4,000

The Mathematics of today's duplication problem:

The Possibility for Duplication If There Is No Coordination Among Researchers:
If the 300 million people in the United States all did their genealogy for 12 generations back using traditional methods, that means 8192 ancestors for each person or 2.6 trillion names in all. Those 2.6 trillion names divided by the 70 million *actual* people who died in the United States before 1930 gives us a **37,000 average duplication rate** for each of the unique names in the database. This is obviously a ridiculously wasteful situation which cries out for an efficient solution.

This new industrial process is almost as easy as creating a nationwide telephone book.
 Can you imagine everyone in the US having to create their own phone book for whole cities and metropolitan areas if they want to call people and businesses in each of those areas? That ridiculously repetitive process is an illustration of what people are trying to do today in traditional genealogy research methods. Starting a few generations back, many thousands of people are descended from the same ancestors. One might expect researchers today to team up in identifying those common ancestors, but that is rarely done in today's genealogy research world. Instead, they all endlessly repeat the same research steps to find the same people. Eliminating that massive duplication of effort, and adding some major positive efficiencies, makes the whole process at least 1000 times more efficient, and everyone benefits.

Convert 18 billion public documents and publications to 0.5 billion finished unique genealogy names

The World's Source Documents Online Include much duplication

LDS Church online offerings:
3 billion records? Free

Ancestry.com online offerings:
Claim 10+ billion records.
Subscription of $300 a year?
Gross income of $400 million a year.

WorldVitalRecords/FamilyLink/
MyHeritage offerings:
Claim 4+ billion records
Subscription of $200 a year?
Gross income of $100 million a year.

Other sources.
1 billion records?
Subscription of $100 a year?
Gross income of $50 million a year.

Totals:
Assume 18 billion records online.
Subscription fees of about $600 per year for them all.
Assume gross income of $550 million a year.
20 years = $11 billion gross income

(box height 18 on left, to box height 5 on right.)

Compress world data into final form using mass production techniques.

Alternate strategies

Current industry strategy calls for endless, highly duplicated work going on essentially forever, with little or no further advancement towards completing all the basic genealogy of the US.

In contrast, the logic of the new system is to greatly accelerate the high quality compilation and sale of all names, collecting all possible profits within perhaps a 10-year time-frame. Instead of $550 million gross/$150 million net a year forever, collect up to $30 billon gross/$15 billion net in ten years. That equals about 100 times the rate of profit earning over the 10-year period.

Professionally assembled individuals and their family relationships: 0.5 billion unique names.

Professionally assemble the individuals and their family relationships contained in the world's supply of "raw" online public records. From the 18 billion record entries, assemble about 0.5 billion unique names in all. Sell pedigree information to individuals for a small fee per name, perhaps $3.
For phase 1, taking 2 or 3 year's time, assemble 70 million names (US residents who died before 1930). Estimate gross sales of $6 billion, with net income of at least $3 billion. Continue more expansion phases until project reaches 500 million names assembled, with up to $30 billion gross sales, and at least $15 billion net income. This compresses all profits from sale of genealogy data into a much shorter time frame.

Convert/reengineer a portion of the genealogy industry to a highly efficient industrial system

Inputs to Genealogy Industry

$66 billion annual inputs to private genealogy industry :

$6 billion cash inputs
$60 billion labor "sweat equity" inputs

Use modern techniques to convert and reengineer a large portion of the nation's genealogy hobbyist and cottage industry professional activities to a highly efficient industrial format.

This is similar to the prepared food industry which has moved large portions of food preparation out of the home and into the factory -- TV dinners, mixes for entrees, salads, soups, fast food restaurants, etc.

Doing genealogy research using industrial methods would make the results have high quality and a low price. That should greatly broaden the market for this data.

Convert portion of hobbyist "sweat equity" into paid industrial work, using mass production techniques:
1% of industry = $0.66 billion per year.
For 10 years = $6.6 billion gross income, with at least $3 billion net income.

Convert just 1% of "sweat equity" to paid industrial work. Much more is possible.

Earn 10 times as much money each year on a shortened project
Graph 1

Alternate strategies

1. Current industry strategy calls for endless, highly duplicated work going on essentially forever, with little or no further advancement towards completing all the basic genealogy of the US.

2. In contrast, the logic of the new system is to greatly accelerate the high quality compilation and sale of all names, collecting most possible profits within about a 10-year time-frame. Instead of today's expected $0.55 billion gross/$0.15 billion net a year for 100 years, collect up to $3 billion gross/ $1.5 billion net per year, or $30 gross/$15 billion net over about ten years. That equals about 6 times the annual rate of gross earnings and 10 times the net earnings over the 10-year period.

$billions

3

2

1
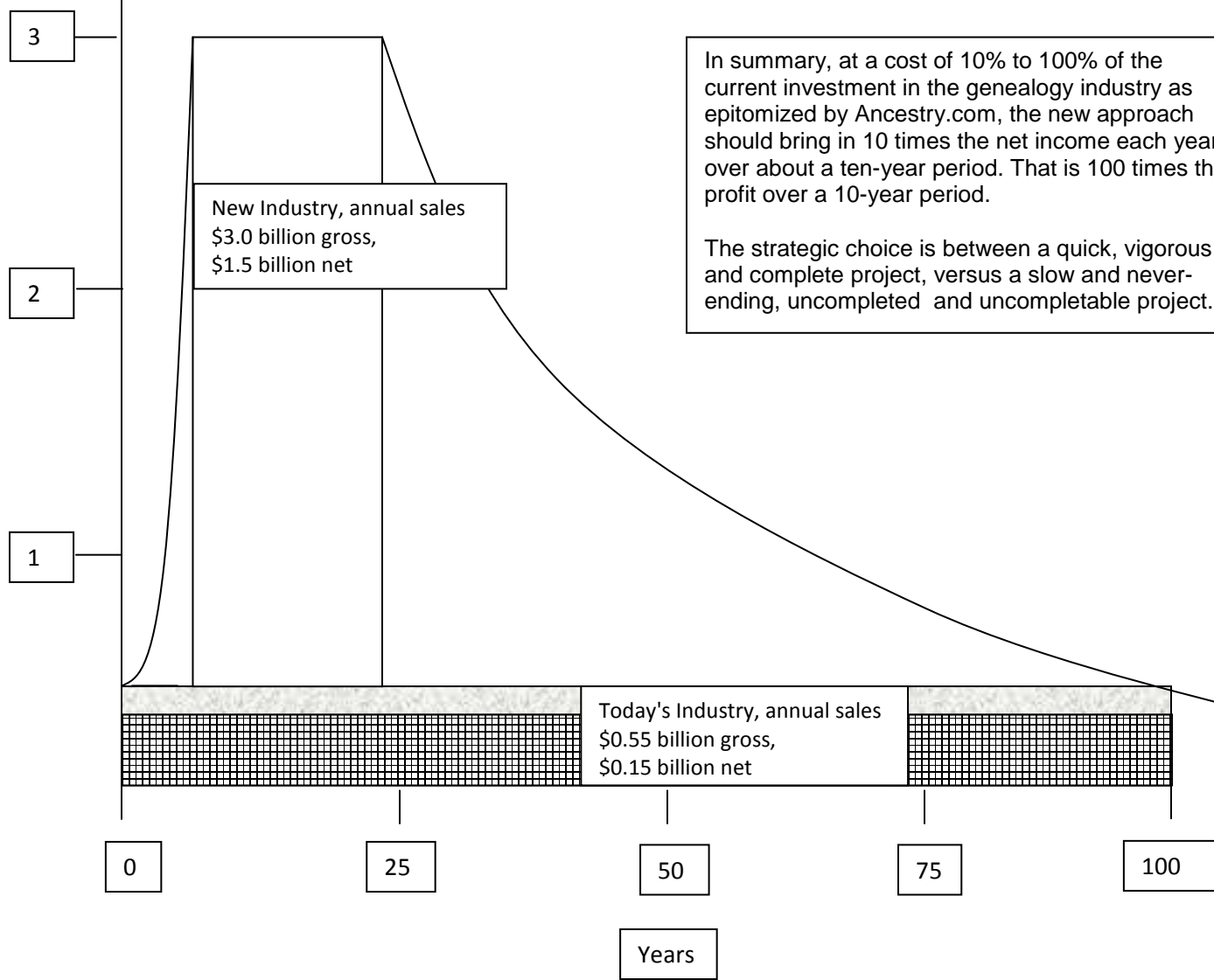
New Industry, annual sales
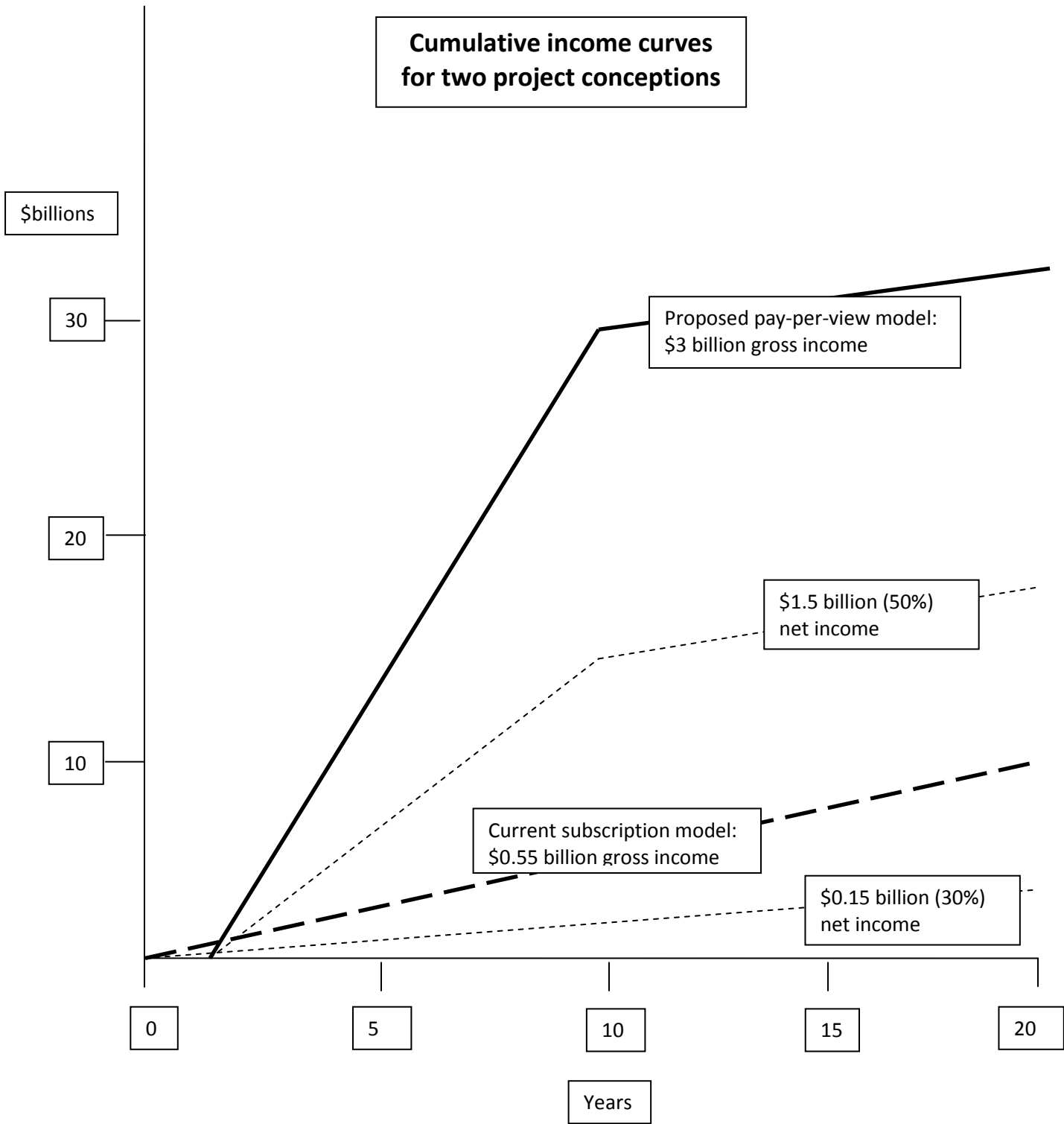$3.0 billion gross,
$1.5 billion net

In summary, at a cost of 10% to 100% of the current investment in the genealogy industry as epitomized by Ancestry.com, the new approach should bring in 10 times the net income each year over about a ten-year period. That is 100 times the profit over a 10-year period.

The strategic choice is between a quick, vigorous and complete project, versus a slow and never-ending, uncompleted and uncompletable project.

Today's Industry, annual sales
$0.55 billion gross,
$0.15 billion net

0   25   50   75   100

Years

Earn 10 times as much money each year on a shortened project
Graph 2

**Cumulative income curves
for two project conceptions**

$billions

30

20

10

Proposed pay-per-view model:
$3 billion gross income

$1.5 billion (50%)
net income

Current subscription model:
$0.55 billion gross income

$0.15 billion (30%)
net income

0   5   10   15   20

Years

Comparative computer processing times

Less than 2 hours of Verizon telephone calls or Walmart sales
would pay the computing costs for assembling all basic US genealogy data

Many people seem to imagine that computing the genealogy for a country is an infinitely large project which even today's supercomputers could never do. After all, we've been doing this for hundreds of years and we have hardly begun to scratch the surface, right?

However, that is a completely inaccurate misconception. If we were to process one name every time Walmart sold one item, it would take 1.75 hours to process all the people who died in United States before 1930. Does that sound like it's too much for us to do?

Or let us look at Verizon telephone company. They are one of the largest telephone companies, with perhaps only AT&T being larger. The Verizon company processes 1.5 billion calls per day. If each one of those calls represented our processing one name of the historical genealogy of the United States, it would take 1.12 hours to complete all basic US genealogy. Does that sound like it is a process beyond our capability?

Calculations:
Walmart item sales: Walmart had a gross income of $347 billion in 2008 on 7.2 billion purchasing transactions. That is about $48.19 per sale. If we assume $1.00 as the average price per item, that would be 48 items. 7.2 billion transactions a year means 20 million a day. That would be 20 million transactions per day*48 items per transactions = 960 million items per day. That is 40 million items an hour. At a similar speed, processing 70 million genealogy names would take 1.75 hours.

Verizon telephone calls: 1.5 billion calls a day or 62.5 million calls an hour. That is only 1.12 hours of computer processing time for 70 million names.

Another part of the solution:
Augment the traditional image collection process
with an Internet-based crowd-sourcing methodology
for the remaining 22 billion world genealogical images.
Personal computing devices, including mobile devices, can extend the
"camera crew" concept to every genealogy hobbyist on the planet.
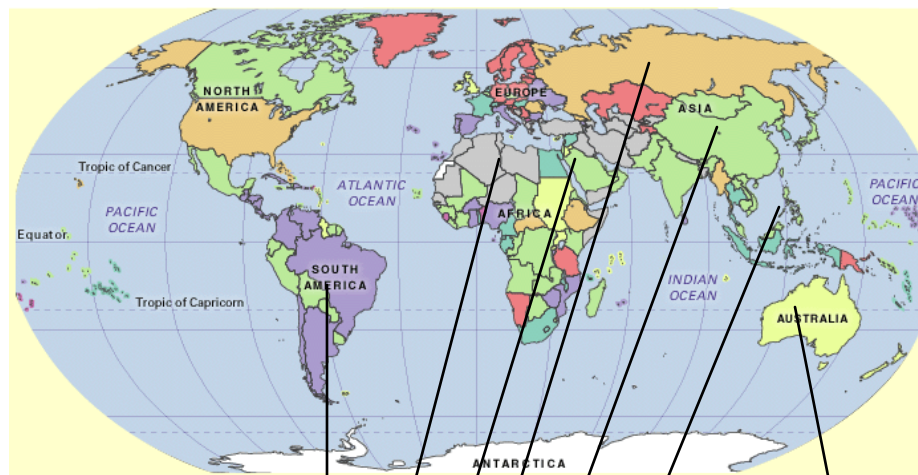
Internet connections.

Assemble remaining 22 billion images, out of 30 billion.

A high quality, verifiable central database, containing no duplication, both depends on and supports a worldwide process of individual genealogists adding names and accompanying documents to that central database. Bypassing the current industrial data collection and preparation "silos," which, although they do create a very high quality product, are limited by the operational budget and management of a single organization. Current data collection processes, even though they are very vigorous at 100 million images per year, will require about 220 years to finish the basic collection on the remaining 22 billion images. An alternate method could get the project finished within a reasonable time, while also saving the LDS Church billions of dollars.

The LDS Church and Ancestry.com are still operating mostly relatively slow and cumbersome (but high quality) "old industrial" institution-to-institution, contract-bound and law-bound record collection and processing methods. This kind of process may help Ancestry.com to retain its control over the data it offers to subscribers, but it is not clear why the Church would resist receiving and storing images and data directly from people around the world.

A new Internet-based crowd-sourcing mechanism, often using mobile devices, could augment or bypass the older methods, including many of the contractual or other legal constraints on institution-to-institution record acquisition processes. Individual citizens in their own country are likely to have freer access to genealogical data than outside religious institutions, there is unlikely to be contractual constraints on the use of that data, etc.

As an example of what can be done today, see the billiongraves.com project, a successful, mostly volunteer-based effort that captures, transcribes, and indexes headstone images from around the world. It uses iPhones and Android phones to photograph and store GPS coordinates for each image. If we knew where all the data was, and had at least a low-quality copy of it, we would then always be able to choose the best places to send the more formal camera teams.

We should assume that if they are not soon captured, most of the remaining 22 billion images will be lost forever before the end of the 220 year wait for formal, industrial imaging. It would be better to have a low-quality image rather than no image at all, although today's 16 megapixel cameras can do a very acceptable job if the lighting is good. The camera crews use 50 megapixel cameras and often position them further from the documents. A 16 megapixel camera closer to the documents might produce as high a resolution image (300 dpi) as a 50 megapixel camera further away.

If 1 million genealogy hobbyists and enthusiasts worldwide each took 1000 images a year, which is about one camera memory chip load, they could complete 1 billion images a year, and theoretically finish the entire job within 22 years. A 22 year operational timeframe certainly sounds better than a 220 year timeframe in this situation.

Many of these images would have immediate value if placed online, so that it should be possible to support this project through subscriptions, so that the LDS Church budgets are not the main limiting factor. Image suppliers could earn some compensation for their images, and even more income for the completed genealogy work.

Providing a data storage facility that allows the capture of all these images is feasible and should have a high-priority.

Is there a problem to be solved in the genealogy industry?

I am proposing a solution to the problems of the LDS Church with its extensive genealogy projects, and, at the same time, I am offering a solution for the problems of the broader genealogy industry, including a new and more profitable business model for investors.

But, wait! Does anyone at the LDS Church or in the broader industry even recognize that they have a problem? And, if there is a problem, does anyone care? There are perhaps 100,000 serious genealogists within the LDS Church, and perhaps 4 million serious genealogists in the United States. Do any of them recognize a problem, and if so, do any of them care about it? Most of the professionals are happy with the current situation and their existing business models, and see no reason to change.

Why does anyone do genealogy research work as a hobbyist or volunteer? Members of the LDS Church have been told they have a religious duty to trace their ancestors. But what does that mean? Is covering four generations – yourself, your parents, your grandparents, and your great-grandparents – enough? If so, that data can often be gathered in outline form with only a few hours of work. Why would anyone want to go back to earlier generations or to add more detail for each generation?

For those outside the Church, the reasons driving hobbyists may be nothing more than curiosity, a desire to know their "roots," or it may stem from a desire to claim membership in some organization such as the Daughters of the American Revolution.

The LDS Church has unofficially accepted the goal of gathering genealogical information* about everyone who ever lived on this earth. Of the 70-90 billion people who have ever lived on this earth, it is estimated that records exist for 10 billion of those people, in the form of about 30 billion record pages or images, which contain, overall, about six different entries for each person. In other words, there are about 60 billion record entries which all need to be brought together and integrated and compressed to describe each of the 10 billion unique people as completely as possible.

> (*Although it would probably be difficult to find a clear and complete public statement on this point, we can easily deduce that broad goal from the fact that the Church is spending about $100 million a year on computer software and database development, while spending perhaps another $200 million or $300 million a year on a genealogy library system supporting about 4000 branch libraries, on genealogy record image acquisition from repositories around the world, and on cataloging, digitizing, and indexing all of those records, etc. Also, many of the Church managers, employees, and volunteers engaged in this extensive activity assume that the corporate goal is to eventually complete the work of gathering and correlating the contents of all existing world genealogical records.)

Does the Church actually have a system for doing that correlation, integration, and compression of the names after all the records are gathered together? The institutional Church would say yes, but the serious and informed member and non-member users would mostly say no. The main lesson to be learned from the Church's newly created databases is that traditional methods of research produce really terrible, almost unusable, centralized results. The massive duplication and disagreement and confusion found in current Church ordinance databases are a great source of discouragement and disappointment among regular members. One genealogy educator opined that one half of the Church genealogists simply refuse to have anything to do with current Church systems, I presume mostly because they are of such low quality and cannot be corrected in any reliable way.

The only difficulty with this broad project conception is that, at current rates of progress, it will take about 220 more years to finish the basic record collection process (about 100 years have been invested already), and it will take essentially an infinite amount of time to extract and integrate all the useful information found in those records, since project progress on the data integration front moves forward at a

barely measurable rate, if at all, while consuming hundreds of millions of member-genealogist labor hours, plus the labor of many non-members.

The outsiders generally don't have as broad a goal as the Church, but they would be happy to accept the benefits of it. But will they also help contribute to it, or will they only expect to receive its results for free? I believe a high-quality, properly conceived system would be gladly welcomed and financially supported.

Can these two large processes, image collection and name-integration, especially the name-integration process, be accomplished in a reasonable time and at a reasonable cost? I am saying yes, and offering a very specific solution involving extensive cooperation among genealogy hobbyists and volunteers.

Social resistance factors
There is a powerful tension between the old ways and the needed new ways. As I mentioned, most of the genealogy professionals, in and out of the church, are perfectly happy with the old ways concerning how names are integrated into finished genealogy work. From their traditional business standpoint, they have no reason to be for, and every reason to be against, increasing the efficiency of assembling names by 1000 times, as this new project proposes. They want to preserve their personal business options.

It seems that the Church's program almost perfectly follows the preferences of current genealogy professionals. Professional genealogists are always happy to have more raw records available in digitized and indexed form so that they can have more research options to offer their clients. Consistent with that preference, the LDS Church is spending hundreds of millions of dollars every year to meet that goal of professional researchers. At the same time, the professional researchers would resist, either consciously and vocally, or only more discreetly and subconsciously, the idea that there is a better way for names to be integrated into final form, especially if that means that their options to acquire new clients could be impaired. Perhaps it is this simple self-serving impulse which keeps the final steps in the record processing, the high-quality assembly and integration of individual names and their associated source documents, from being implemented in a highly efficient way. This seems to put some of the Church employees and volunteers in position to strongly resist the larger goals that would benefit everyone.

If the Church's genealogical operations were not in the hands of the professional genealogists, it is likely that some time in the last 13 years, since the first version of FamilySearch was initiated in 1999, the suggestions for how to do these processes far more efficiently would already have been accepted. As it is, it appears that the resistance levels are actually going up as the ideas for improvement are becoming more concrete and better explained. Apparently that restrictive and even monopolistic guild mentality will have to be addressed directly before the problem can ever be solved.

One solution to this self-protective guild mentality, perhaps the *only* practical solution, is to simply take the issue out of the hands of the professional genealogists inside and outside the Church, by reorganizing a portion of the genealogy industry along the new lines which are suggested here. That obviously would take a fairly large amount of further investment capital. Gradually increasing increments of $20,000, $200,000, $2 million, etc., invested could quickly show the power of the proposed new methods, plus demonstrate the profitability of the new business model.

The large private company Ancestry.com (and perhaps other similar companies) is also perfectly organized to keep the genealogy professionals happy (and the hobbyists isolated and relatively ineffective). They make the raw data available online, but only supply minimal abilities for nonprofessionals to cooperate in highly efficient name assembly processes. Ancestry.com apparently makes a very large portion of its income from unproductive duplication processes. For example, if there are a hundred people all working on the same genealogy research problem, mostly unknown to each other, then Ancestry.com can collect all of their $300 annual subscriptions. If that group were coordinated in some way, they might instead need only one or two subscriptions to Ancestry's service. This professional resistance to serious hobbyist and volunteer cooperation, both inside and outside the Church, is almost certainly the largest barrier to establishing appropriate technical solutions.

Reengineering the Genealogy Industry
for Genealogical Success, Religious Fulfillment, (and Profit?)

Introduction

1. For the concerned LDS Church member, we want to make clear the reason why this project is so important, valuable, and useful. One might argue that the entire point of the LDS Church computer software and database efforts should be to create a high-quality database in which all historical ancestors of LDS people are identified and linked to all their other family members. Without this final step of linking all the people together into families, the very important step of sealing families together into patriarchal lines cannot be accomplished.

Since 1938, the LDS Church has put enormous resources into microfilming and centralizing the world's genealogical records, perhaps the equivalent of $100 billion in today's currency. In more recent times it has added a vigorous transcription and indexing process so that all of these public records can be quickly accessed. What is still missing is a highly efficient way to assemble names from these many public records into a finished genealogical database which covers whole nations.

2. For the investor or planner, we will try to show the size and shape of the overall project and what else is needed for success. The current business organizations and databases and business models assume that the process of individual researchers piecing together their pedigrees will go on essentially forever, with barely any measurable change each year. All those companies which offer online access to billions of public records are taking in perhaps $550 billion year, with perhaps $150 million annual net profit. Under current conditions and using current methods, that process will likely go on forever at a certain relatively low fixed rate.

Under the proposed new methods, new business plan, and new business model, the process of assembling high-quality names can be done up to 1000 times faster than with current methods. This means that the net profit from the genealogy industry can be 10 times larger each year than it is today, or 100 times larger over a 10-year period. That will likely mean that the project will be greatly compressed, so that it operates vigorously for perhaps a 10 year period, and then slopes off to a lower annual level.

Obviously, it will be necessary to find people who are concerned enough about genealogical issues to take the time to study this presentation before anything substantive can happen. Hopefully, this high-level overview, plus other more detailed documents about the processes and tools, will be enough to show the need and feasibility of this proposed project.

20120828 the names story, in and outV04

Implementing The Solution

After all the charts and graphs, how do we integrate all of this information into an understandable business story?

Getting the names all together is probably many times easier than most people imagine. The point is, we are focusing on bulk collection of names, usually doing one surname at a time. There are boatloads of names in this form on the web and in published books. For example, the Huff book contains about 15,000 names, including 5000 of the Huff-surname descendents of Engelbert Huff, born 1637. That needs to be converted into complete electronic form and placed in the appropriate part of the database. On the web we have the Leavitt family which is has gone to great lengths to collect about 165,000 names, including about 15,000 with the Leavitt surname. Something similar can be found online for the Brigham Young family that was perhaps only 1000 Young surname descendents.

There are at least 60,000 books in the LDS genealogy Library, many of them digitized, which are in the form of "the descendents of...". These names can be quickly entered into our database and then, with an extra partially automated step, the names can be checked against the records in the Ancestry.com database so that source records can be found and associated with all these names. My hope is that for this first effort of collecting these names, the cost per name can be as low as one or two dollars per name.

This process turns the Ancestry.com database into a gold mine, as we assemble names quickly either from these prepared sources I have mentioned or extracted in surname groups from the Ancestry.com database (or the Church database), and in either case the names are either associated with a source record, or can be quickly associated with a source record.

There is probably up to about 500 million unique names which can be plucked out of the perhaps 8 billion raw names which are available through Ancestry.com. The main question then is how far does one go in building the initial database before the sales began.

Once we have the initial database of perhaps 300,000 up to 3 million names, then we have to find customers for those names.
There are many potential categories of customers, and I plan to go quickly through some of those kinds of customers and guess what their interests might be in buying these names.

The LDS Church -- At this point, they have an unlimited supply of low-quality names to take care of their obligation to supply names for temple worship. This probably means that they will have very little interest in this project at the beginning.

LDS members -- They also do not have much incentive to improve the system to get better data, especially if it costs money, simply because the Church has found a workaround which allows them to supply names for everyone with almost no noticeable effort. The Online Indexing system produces at least one million names a day, so they can get all the names they need for year with about one week's work.

LDS member genealogists -- There is a small group among the LDS Church members who pay more attention and will likely be more interested in this new project. In one fairly typical case, a woman who is the director of an LDS genealogy Library has spent her life doing genealogy research and has completed

six full generations plus has taken two lines back to 10 generations. So this possibly 10 years of work on her part has yielded about 158 names that are her progenitors. She has done far better than most, but that is still a pitifully small number of names compared to what the records would hold concerning her progenitors.

For her, it might be well worth the $200 it would cost for her first five generations of ancestors. This would be a chance to check her work and possibly find new data of interest to her. She apparently has some interest in going back 10 generations. If she were to order the entire 10 generations pedigree, at the deeply discounted price of $2000 (it should be about $6,000), then she would have an unbelievable gold mine of information. She would have 1024 surnames and 2048 people, easily more than 10 times the work that she did in her 10 years of labor.

So perhaps 100,000 church members would have the enthusiasm necessary to participate fully. 100,000 members times $2000 equals $200 million. That would be a tidy sum and would easily justify doing this project.

Let me mention how one could go about building this initial database. It would make sense to begin with the core of Utah genealogy, all the pioneers who came here and their descendents. There should be many sources of data for these names. That is a great place to begin a database and to branch out from there, but because these names have been researched and sent to the temple so many different times, their market value may not be very large. We may have to get beyond this pioneer ancestry group before the members will have much interest in our work.

It's interesting to note that the University of Utah has an 8 million name database of Utah families and their family relationships. It is used for some of their medical and genetic studies. They take this structure of names and add in medical information from various sources and then are able to use it in their research. So one might guess that we would have to get up into the 8 million name range to generate a lot of interest. At that point we could offer some of this information for scientific purposes, offering to compete with the University Utah/Huntsman Institute complex and their database of names.

We might wonder whether we could make it clear to church members that the church is satisfied with using these low-quality names, and that they do not contain enough information to do sealings. Could we embarrass the church into raising its standards and buy fully researched names from us, or could we embarrass the members and encourage them to buy their names from us which they submit to the temples? It's hard to say. It seems more likely that the Church members would get the message, and buy names from the system and send them the temple. They have now seen the mess which is the FamilySearch database, and they might be happy to purchase clean information to send to the temples.

If the church were in fact willing to upgrade its standards, and pay us $10 for each high-quality name, that would mean a cost of $50 million a year for the 5 million names which they need. If they were willing to pay that money, that would allow us to supply the names they need and also to move on to do a much larger number of names.

Genealogists outside the Church -- There are probably 40 times as many US genealogists outside the church as there are inside the church, so that would be a very important group to target. As I calculated elsewhere, this group of people is spending about $66 billion a year in time and money in the United States to pursue their roots. It seems likely that many of them would be willing to pay a few hundred dollars for that convenience of getting high quality prepared data. If all 4 million of the most active genealogists were to pay $200 for their five generation pedigree, that would give us $800 million in income. As I mentioned before, we seem to have all the data available we need from books, the Internet, and Ancestry.com, to compile all the names to meet this need. At a cost of one dollar a name, or an eventual cost of $70 million in all, with a return of $800 million, that would be a reasonably good business proposition. If we sold the highest quality names which cost us $10 each to prepare, that would be $700 million in costs, but the data could be worth up to $8 billion over a 10 year period, if these genealogists were sufficiently excited about the product.

There is one possibility of delaying the costs until we have actually sold the data. Once we have demonstrated how the system works, it is possible that many genealogists will be willing to enter their high-quality data on consignment, and when we sell that data to the public, we can transfer royalties to the original data suppliers. If it works, that would help us avoid some of the enormous upfront costs if we were to build the database nearly completely before we began sales.

The man in the street – I am also assuming that there will be millions of people who have never had the time to get involved with genealogy research work, but still have a sincere interest in learning about their progenitors, their roots. For example, if one person in 20, 5% of the populace, were willing to pay $200 for a five generation pedigree, that would mean 15 million people paying $200, which is $3 billion.

Of course, there's no way to predict in advance which of these various sources of income will work out, but with the huge national interest in genealogical data, as evidenced by the $66 billion estimate of yearly inputs to the genealogy industry, and the extreme activity on Google and all of the specialized sites, surely some combination of these groups of potential customers will be willing to support our project.

20120824 Implementation of the genealogy solution-trim

A Simplified "Problem and Solution" Treatment of the Situation

The Problem

1. Bottom-up genealogical research which is immensely inefficient.

> The inability to coordinate efforts among users, and the resulting overwhelming levels of time-wasting duplication of research, make it mathematically impossible to ever finish any significant amount of the nation's basic genealogy research.

2. Church, and professional, and non-Church amateur genealogists who are set in their ways, if not actively resistant to a new approach.

> Professional genealogists are happy to have unlimited access to images of the records of the world. However, in general, they want no serious competition for their paid work of assembling high quality family structures. Under current methods, their work requires much education and high skill levels, much like the master craftsmen of old, and they resist the "mass production" methods, with their much lower knowledge and skill requirements, even though that would solve the Church genealogy data processing problem (and the problems for all amateur genealogists.) Church professional genealogists consistently choose personal business success over a completely successful Church genealogy and temple program. Non-Church genealogy professionals at least only see the conflicts between the professionals and the hobbyists, without the added dimension of the needs of the Church.

3. Professional genealogical database system operators like Ancestry.com are exacerbating the problem by churning in order to sell services for money.

> If their clients participate in massive duplication of research efforts and the repetitive posting of the confusing results, then Ancestry.com maximizes its subscription income.

4. Confusion in the Church temple genealogical system allowing members to change data and thus creating multiple copies of bad data.

> This is a consequence of the communal access system which was chosen for the current main system – users can make changes which affect tens of thousands of other users. The current methods work well for the (intended) narrow functions of submitting names for temple work, and tracking that work's progress, but is not suitable for entering the original family structures. (This can be remedied simply by giving each computer system user their own private space which no one else can change.)

5. Mullions of names are processed through the temples even though they are unlinked to the names of other family members.

> This is a natural consequence of relying on the various "name extraction" methods the Church has used over the years, which results in long lists of unconnected names. The number of names needed for temple ordinances far exceeds the ability of members to produce fully family-linked names using current methods. (This decades-old problem can be remedied fairly easily with the new system.)

The Solution

All genealogists will find the next two steps very useful:

1. Build a clean database by working top-down and by using original sources.

2. Link the families.

> When looking at public genealogy databases today, it is easy to see why it is a revolutionary thought simply to attempt to build a high-quality, clean genealogy database without duplicates.

No one has such a product, or has even seriously attempted it. And in most cases, if data is high-quality as it exists on a researcher's PC, it will be very seriously degraded when it passes into any central site. This must be remedied.

Of course, building this clean database is the heart of the solution, and nearly all the special manual procedures and special programming features are focused on reaching this solution. The patent which describes this procedure is very lengthy and detailed, but it can be summarized by mentioning four of the most important features:

1) Collecting and entering data in descendent sequence, beginning with an ancient ancestor, provides previously unsuspected efficiencies at several levels, and makes very large "mass production" efficiencies a reality. Data can be collected and entered up to 30 times faster from public records, since all data about a particular surname can be quickly extracted from whole record sets and organized into families, and then those finished family surname structures can be quickly interconnected through marriages. This process can improve individual researcher productivity hundreds of times.

2) A unique number is assigned to every name submitted to the database, with that number identifying the owner of the data submitted and the surname structure to which it belongs. This makes possible connecting together in one place all known data about each unique person, wherever that data may reside on the Internet. It also makes possible the identifying and avoiding or suppressing duplication among system users.

3) The quality of most online data is abysmal and is a major cause of enormous duplication as researchers have to recheck every name they choose to use. In the new database, it is relatively easy to segregate the high-quality data from the low-quality data so that system users will not waste their time with low-quality data during such processes as linking families together through marriages.

4) The lack of fairness among participants in today's mostly hobbyist and volunteer genealogy research processes is often a source of irritation and even bitterness. People may do volumes of high-quality work, only to see it go to other people without any compensation whatsoever, in recognition, shared data, or money. It is often especially irritating to see one's work appear in a mangled form in one or more public locations, with another person claiming authorship. In the new system, data owners can be clearly identified, with their reputations enhanced by their high-quality work. In a full implementation of the system, the owners of high-quality data can at least be granted "points" as others use their data, and those points can in turn be expended in taking advantage of the work of others. A user's net number of points can be turned into money in an ideal implementation, so that those who do the most high quality work get the most reward. At that point, it becomes feasible to have large numbers of semi-professional researchers earn their livings doing large amounts of high-quality work.

Only LDS Church genealogists are likely to care about the next four points:
The current LDS Church Temple systems were designed for these exact processes and should work perfectly well for all four points. The new system probably only needs to supply a means of transferring data to the Church temple systems.

3. Vet all the names to ensure temple work is done or not done
4. Prepare the cards for temple work names
5. Submit the names to the temple
6. Document that the new temple work is done