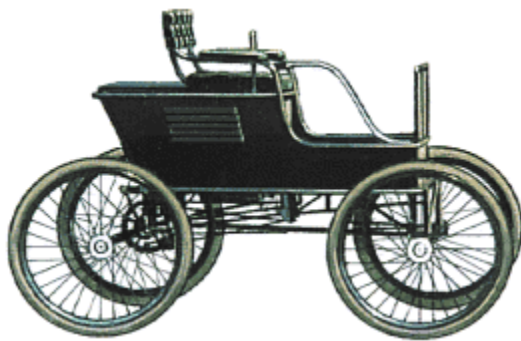


Doing Genealogy the Henry Ford Way

Assembling High Quality Genealogy Data
2000 Times Faster
Using Specialization and Cooperation



The first Stanley Steamer (1897, United States) was a steam-powered, chain-driven buggy with a maximum speed of 25 mph.

Traditional genealogy research methods have been improved somewhat by computers, but to only 5% of what is now possible

The new cooperative genealogy research methods are 2000 times faster than traditional methods



2011 Lincoln Town Car (from Ford)

Doing Genealogy the Henry Ford Way

**Assembling High Quality Genealogy Data
2000 Times Faster
Using Specialization and Cooperation**

Preface

Everyone who does genealogy research realizes there are many problems with current methods (which I will also call traditional methods since little has changed in basic concepts during my 70-year lifetime.) Unfortunately, these genealogy problems have steadfastly resisted effective solutions. Over the past two or three decades, hundreds of millions of hours of thought and labor, and hundreds of millions of dollars in capital have been invested in designing and creating new computer systems and new database systems, but the results have been quite disappointing. One could argue that in some areas things have gotten worse instead of better. For example, in many cases, the extra computing power has simply made it possible to generate duplicates at a faster rate.

Many people probably agree that the entire genealogy industry needs to be reengineered to take full advantage of today's technology, but for some reason that has not happened, at least not on the scale it should have. From my experience in grappling with these problems, the difficulty seems to be that if a person looks at only one problem at a time, it is essentially impossible to solve the overall constellation of problems and maximize the benefit of the reengineering project. Thus, for example, as we have seen, if one is upset with the massive levels of duplication which have occurred in the past, one might be willing to spend many tens of millions of dollars directly attacking the duplication problem. However, upon consideration of the much broader set of problems, it turns out that "duplication" is not really a unique problem, but rather a symptom of several other problems, and attacking that symptom by itself does little or no good. One must identify and examine the entire list of problems, all at once, before the optimal solution can be found.

The basic process of this redesign of the genealogy industry was to seek out every problem which could be identified, put them all in one list, and then seek practical computer-based solutions to each and all of them. Separating the causes from the symptoms was one very important part of this analysis.

Table of contents

1. Introduction to the Mathematics of Genealogy Productivity -- Comparing Old and New Methods. p.4
2. Solving The Six Biggest Problems in Today's Genealogy World. p.7
3. The Hidden Benefits and Cooperation Power of Descendent-Sequence Research. p.16
4. Cottage Industry Versus Industrial Revolution Methods. p.21
5. Other Important Features. p.23
6. Brochures used at genealogy conferences, including a short discussion of the most important concepts and solutions. p.26
7. A more complete discussion of the solutions offered and other opportunities found to help improve efficiency and overall value. This was the input to the patent-writing process. p.36

1. Introduction

to the Mathematics of Genealogy Productivity
-- Comparing Old and New Methods

I realize that many people feel allergic to charts and numbers, but the full impact of this new system cannot be understood without considering a few numbers.

Why do we need a new genealogy industry research concept, methodology, and tool? What is wrong with the old methods? Let us start out with the well-known questions of duplication and cooperation:

By my calculations, the expected duplication rate for serious and extensive genealogy research in the United States using traditional methods is about 12,000 times. See calculations below. That obviously is a huge impediment to finishing large quantities of high-quality genealogy research. But we now have the option to bring that duplication rate down to zero using a simple procedural change, in conjunction with the appropriate computer support. We can then go on to achieve a further 2,000 times improvement in researcher productivity using the new computer system design which emphasizes specialization and cooperation.

Duplication calculations: If every living person in the United States were to do the research necessary to compile a complete 12-generation pedigree for themselves, that would mean each person would need to find 8192 ancestors. If there are 320 million people in the United States, 8192 times that number equals 2.621 trillion names to be researched. Of course, the total number of people who died in the United States up to about 1930 is only about 70 million people. So, 2.621 trillion names divided by 70 million names gives us 37,428 as the raw duplication factor. If we further assume that, on average, about three living people (who are siblings) have an identical pedigree, that gives us $37,428 / 3 = 12,476$ as the expected duplication factor when creating all unique pedigrees in the traditional manner.

If procedural changes are made within the genealogy industry to take advantage of these huge potential productivity increases of 12,000 times plus 2,000 times, as just mentioned, then it becomes theoretically possible to finish the basic genealogy for the entire United States within a remarkably short period. Here is an illustrative computation:

Computation for maximum potential cooperation: Assume we have 4 million active genealogists in the United States who would be willing to participate in this project. Assume we have about 70 million names to complete to bring us up to 1930. Notice that 70 million names divided by 4 million participants means that each participant only needs to complete about 18 names. That would mean that if researchers spend about four hours on each name, it would take only about *two weeks* work for each researcher to finish their part of the project. As a practical matter, it would probably be easier to work with a smaller group of participants over a longer period of time, simply because of communication and training issues, but the enormous efficiency factors would still apply.

Of course, participants in such a cooperative project would need a new concept and a new Internet tool to allow them to cooperate on that scale and with that precision. That is what the Genealogy Registry System provides.

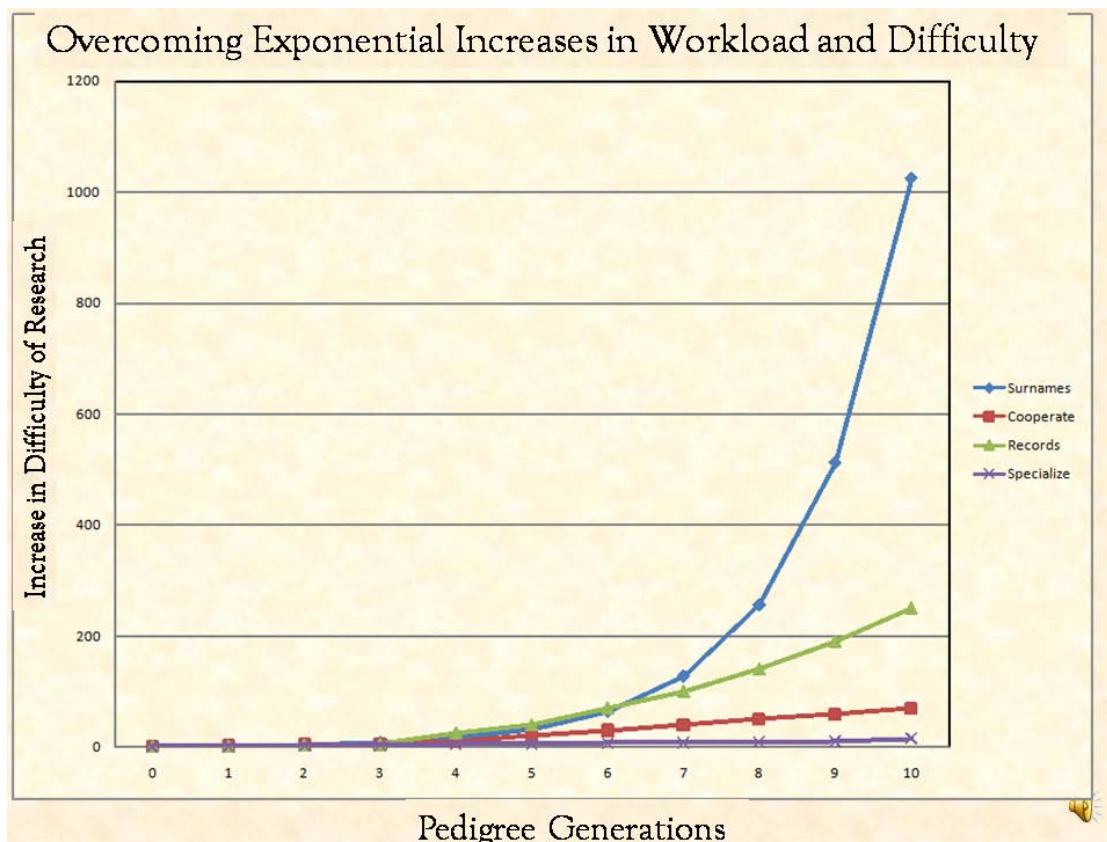
Comparing The Old And New Methods

The difficulties of traditional genealogy research

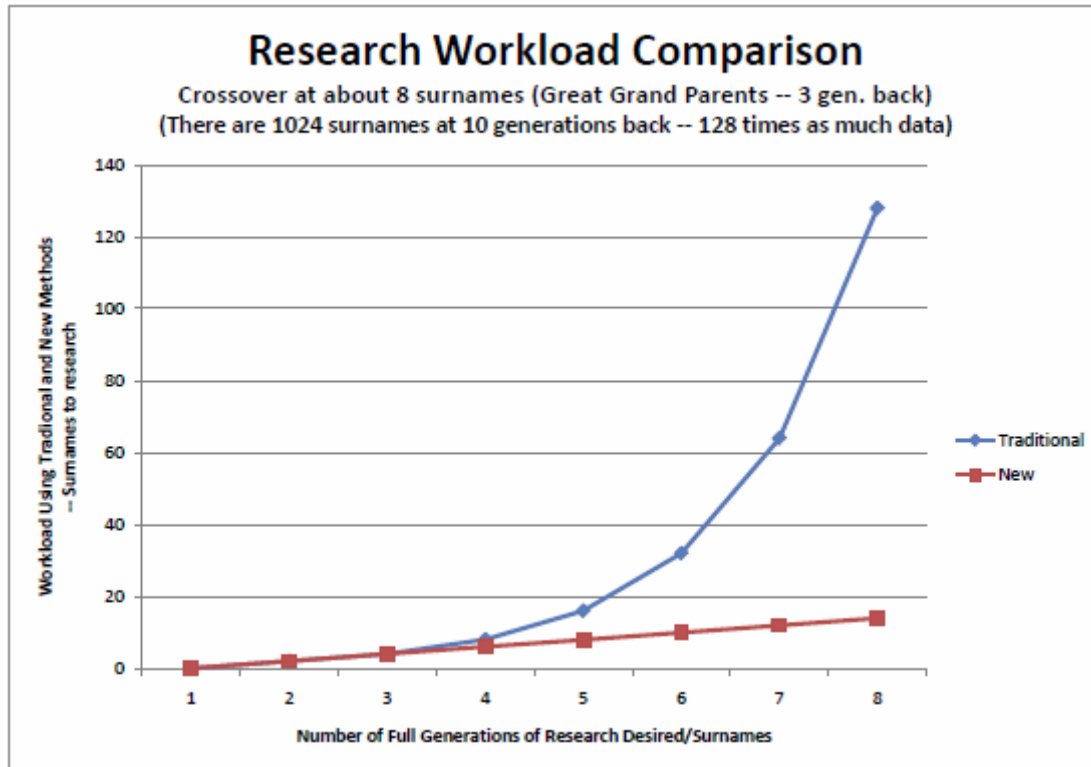
The following chart is intended to illustrate the obvious fact that as you go back in time doing genealogy research, each succeeding generation contains twice as many ancestors, and therefore is at least twice as hard to complete. Ten generations back, there are 1024 ancestors, with 1024 surnames, who might have been born anywhere on the planet. This makes it essentially impossible for the researcher working alone to ever complete a full 10-generation pedigree in one lifetime.

Through cooperation, the new Genealogy Registry system overcomes this vast difficulty to the individual researcher by essentially flattening out this steep difficulty curve, represented by the top blue "Surnames" line, to one which is almost horizontal, as represented by the lower red "Cooperate" line. (This will be discussed in more detail later in the book.) The cooperative "parallel processing" of the new system makes all the different generations of ancestors almost equal in difficulty to complete.

The second kind of difficulty which gets worse as one moves back in time is the difficulty of reading and interpreting the older records. Some of those difficulties of interpreting older records can be overcome if researchers will specialize in those records so that their accumulated professional knowledge can help achieve the highest possible accuracy in interpreting them.



If we enlarge a section of this basic chart to make it easier to read the beginning portions of the diagram, we can see the individual workload comparison between doing research using traditional methods, and doing research using the new cooperative methods. It becomes evident that about three generations back from the researcher, at what we might call the "cross over point," it starts to become noticeably easier to use the cooperation method to gain the genealogy information desired. And as one moves further back in time, the difference in individual workload between the two methods becomes more extreme.

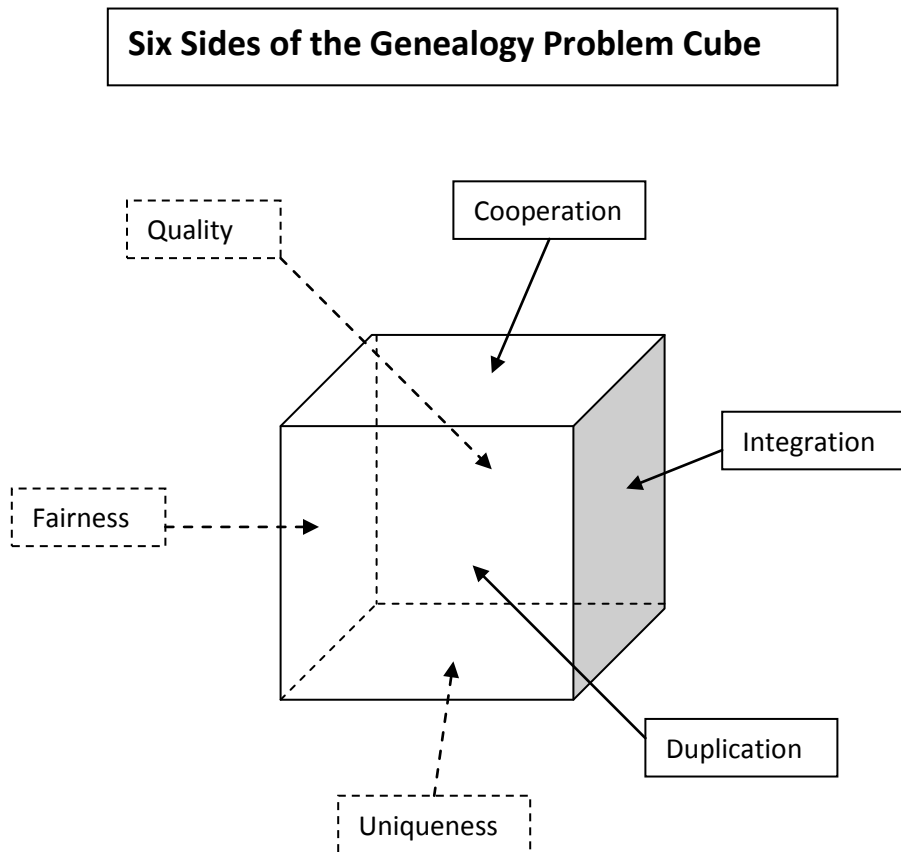


We might notice the bad news that nearly all of the cooperation, sharing, and efficiency problems we have in genealogy research arise simply because we insist on continuing to use traditional pedigree-sequence research methods almost exclusively. Where once that was the only practical option we had for assembling genealogy, it has now become the least efficient and most problem-filled of the available methods, at least in the situation where one wishes to go back to more than three generations of ancestors.

However, there is now plenty of good news since the Genealogy Registry System provides a solution for every one of those very troublesome problems concerning inadequate cooperation, sharing, and efficiency.

2. The Six Biggest Problems in Today's Genealogy World

There are at least six major problems in today's genealogy world (and many smaller problems). Genealogy researchers are certainly aware of these problems, but it appears that no one has ever tried to show all the interrelationships among all the problems so that they could all be solved at once. The LDS Church is to be commended for trying to solve at least one of these problems. It decided a few years ago that the amount of duplication within its own genealogy data was unacceptable, and it has taken vigorous action to solve that problem. From my viewpoint, its efforts have been a disappointment, not from lack of trying hard enough, but simply because they are attacking a symptom of something else rather than the actual problem itself. If we were going to try to solve two problems at a time, rather than just one, we might say, for example, that poor data quality and excessive duplication were two sides of the same coin. But, with at least six different major problems to solve at once, we would need at least a six-sided coin. That doesn't work very well as a metaphor, so I will use a cube which actually has six sides.



Duplication

There are several kinds of duplication, and they occur for several kinds of reasons. We need to list and discuss each kind and its reasons separately.

1. People are unable to coordinate and cooperate effectively, so they repeat large amounts of research.

The first question that people ask as they begin research work is "What has already been done?" But today, there is no reliable way to find out what everyone else in the world has done. That means that the only course open to the researcher, in most cases, is to just begin as though nothing had been done before. I am told that there is one case in LDS Church records where one man's genealogy work has been done at least 10,000 times. That could easily happen, for example, if each of his descendants about six or seven generations down from him all did the same research work, simply because they did not know each other intimately and had no way to know exactly what work had been done before, and what had not. "Doing it again just for good measure" could easily result in the same process being done 10,000 times. Calculation: If each child has 4 children, for 6 generations we have $4^{**6} = 4096$ descendants, and for 7 generations we have $4^{**7} = 16384$. The 10,000 figure lies in the middle.

This kind of duplication means that thousands of people might each spend thousands of hours all doing exactly the same thing, leading to extremely wasteful results. It may be a good experience for each of these people to explore their ancestors, so it is not wasted in that sense, but it is wasted in the sense that they could have done many other good and useful things that relate to their family's genealogy if they were not merely repeating the work of other people.

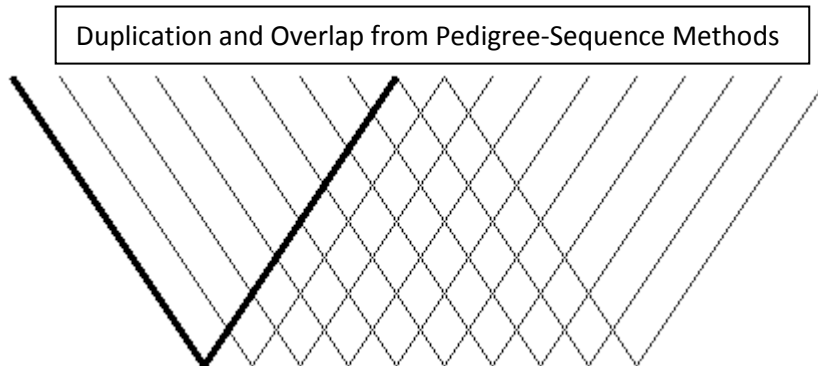
2. When the final research data from these many different duplicate efforts are all compiled together in one place, it may lead to a great deal of confusion for anyone attempting to use that data. Whether all these voluminous, repetitive results are added to a single database or are scattered throughout many databases on the Internet, for example, any later researchers wishing to use this data start out in immediate overload. They have to choose whether to sift through these potentially tens of thousands of differing results to find something which may be of use to them, or the apparent magnitude of that sifting task may convince them that they should simply ignore that data and begin again.

The other factor typically operating here is that the data they find is of very low quality. That is, it may contain a minimal number of the important data fields such as birth date and place, and death date and place. In fact, the entry for a person will often contain only a name and a relationship. And it might not contain any references to source documents from which its data may have been collected. Genealogists often say that data without source references is a myth, or at least one must assume that it potentially is a myth and it must be reverified.

3. Strangely enough, doing extensive genealogy research in the traditional manner requires a massive duplication of education and information for those who participate. As a pedigree-sequence researcher begins his work, every jump backwards in time along any particular surname line may land the searcher in a new set of records, in a new place, perhaps in a new language, potentially anywhere in the world. If every researcher must learn everything necessary to work in that new time and place, all for the purpose of plucking out one or two names from the whole of it, that can mean a truly staggering educational overhead for each researcher. In contrast, an efficient specialization and cooperation system minimizes the amount of new information which researchers need to gain to be effective and successful. Rather than trying to know everything that might be needed to be successful doing American and European research, the specialist would only need to know about one or a few record sets in a few areas, all in the same language. Simply having most researchers specialize, to benefit themselves and all the other people participating in the project, could remove enormous time constraints from the overall process. Certainly, many people enjoy this "eternal learning" aspect of genealogy research. At the same time, however, it helps make the entire process very inefficient from an overall standpoint.

4. The highly duplicative results of pedigree-sequence research are probably known to most genealogy researchers today, along with the usually chaotic results when attempts are made to use computers to

match and merge the potentially thousands of entries for a single name. The typical situation in the pedigree-sequence research world is shown in the following diagram. Notice that with the Genealogy Registry System this kind of duplication is never created in the first place, so there is never a need for this matching and merging process which itself can introduce so much error and confusion.



Summary

Notice that duplication is not itself the problem. Rather, it is actually a result of poor coordination and cooperation, and of the poor quality of data. One cannot fix the duplication problem, once and for all, unless the coordination, cooperation, and quality issues have been solved.

The Genealogy Registry System has several mechanisms for either preventing duplication or for hiding low-quality duplicate records from public view.

Quality

The quality of data found in most public genealogy data sharing places is usually of such low quality as to be almost useless to the next researcher examining it. Because of the great difficulty and time requirements of finding specific names, it appears that most researchers do as little as possible in accurately recording the results they find that might be useful to others, presumably so that they can quickly get on to the next names they are seeking. Therefore they do a poor job of documenting each name they find, almost ensuring that the next person to view their work will find it of only minimal value. The original researcher may see enough in the records to satisfy their own accuracy standards, but does not preserve that proof for others.

In the past, it appears that people had little or no incentive to leave behind high-quality research work because, since we have had no reliable cooperation and coordination system, it is unlikely that any particular researcher doing research work would ever benefit from the work done by anyone else. We might call this a Catch-22 or non-virtuous circle. Obviously, if everyone did high-quality work, most of the duplication and waste could stop, if we also had a reliable place to exchange this data. But of course we have had no such reliable place to exchange data, so these negative incentives continue to control how data is created.

There are many great benefits from having mostly high-quality data produced by researchers, even if it takes them a little extra time to complete. Not only can each subsequent user be confident that the data they are receiving is correct, since they can quickly reverify it using the source records given, but, with a

large number of public source records linked to a particular name, the researcher can then usually learn a great deal more about a particular historical individual. For example if all relevant census records, court records, land records, journals, birth records, death records, marriage records, etc., are linked to a particular person, then it might be possible to construct an interesting sketch of that person's life. That sort of thing is almost impossible today using data available publicly, since the data recorded by researchers is usually minimal and might easily be wrong. Genealogists are typically very skeptical of this kind of data, as well they should be.

Notice that if everyone created high-quality data and made it available to the public, most of the duplication problems could disappear, assuming that high-quality data could be found reliably in a predictable place by subsequent researchers.

The Genealogy Registry System enforces quality standards in several different ways. Data which does not pass the tests for the completion of data fields, or the tests for supplying sufficient source record links, will be hidden from the public until those deficiencies are overcome. Only the best data is actually available for viewing by the public.

Cooperation

Almost everyone seems to sense that our cooperation systems are inadequate. The Internet has allowed genealogy researchers to make some progress in their cooperation. For example, the Cyndislist.com site, which categorizes and indexes perhaps 260,000 genealogy websites, became an explosive overnight success, simply because there was no other way for people to share their work. People can now put their research on their own website, and through the magic of the Google search engine, they can find out what other people have placed online. But although this is a great improvement over the past, it is still very chaotic. If one enters a name one is researching, and gets back 20,000 hits, that may or may not be very helpful. Who really wants to take the time to sift through 20,000 individual online entries to see if one of them might be of interest to them? One could potentially spend hundreds of hours on that one set of possible leads, only to find that none of them represent the person originally sought. There are about 2.3 million people with the Smith surname in US census records. Finding the right person among so many Smiths with similar names would be particularly hard to do.

Where today a researcher might find 20,000 hits on entries made by 20,000 researchers, and wonder whether any of them are accurate or useful, it would be far better to have one entry in one place which had been checked by 20,000 researchers for accuracy. The Genealogy Registry System is designed to turn these numbers around and gain that huge boost in accuracy.

I believe that much genealogy research work is conducted by people finding other people's e-mail addresses and asking them questions. But that is a slow, uncertain, and laborious process for both the person seeking to ask questions, and the person seeking to answer them. It is certainly better than using surface mail to try to accomplish the same thing, but it is still operating 1000 times slower than is possible with a properly constructed system.

With the new system, it would be rare for people to need to e-mail each other. All participants simply put their best data into a system, which automatically enforces quality standards and eliminates nearly all duplication. This means that the best place for a researcher to look is at this online source which contains all the best data which is available. The time-consuming and error-prone e-mail question-and-answer game should mostly end.

Notice how cooperation interacts with quality and duplication.

Just one feature of the new Genealogy Registry System provides a 1000 times improvement in cooperation mechanisms. The participants can enter one unit of data that relates to their surname and receive back 1000 similar units of data that relate to the rest of the surnames they need to construct a full 10-generation pedigree. There are several other cooperation features and more generalized efficiency features which will be described later.

Uniqueness

Obviously, everyone wants to have their genealogy research results be such that every one of their ancestors appears once and only once in their records and in the world's records. That means the end of duplication. But how does one get to a point where every person appears only once?

One very helpful way would be to assign a unique number for each person who was added to the database worldwide. With high quality data available for that person, and all those closely related to him, then it should be fairly easy to establish a unique number which can be used worldwide.

The most likely source of duplicates is easily stopped. It is indeed possible for numerous people to add the same person to their workspaces in the new database, and in each case a number would be generated for that person. To resolve this potential duplicate number problem, the system supports a competitive process for arriving at a unique number for each individual in the database. The basic rule is that the occurrence of a name which has the highest quality data, including being linked into the largest available descendent tree of data and having the most source record links, would be the default number to identify that person worldwide. If people do indeed adopt that number, and add more and better data using that number, then it becomes even more differentiated from the weaker candidates, until the weaker candidates can safely be hidden from public view to avoid any future confusion.

One of the most powerful consequences of having a unique number for each person is that data found anywhere in the world can be linked to that name in the central database. The entire Internet can then be treated as one integrated genealogy database. A family member might place pictures or journals or videos on a family website which could then be linked to the central index site. This would mean that the central site would not have to store huge amounts of data, but would merely need to supply the links which would point a researcher to these family treasure troves.

A unique number would thus foster the elimination of duplication, greatly assist in raising the quality of the data, and provide a powerful cooperation mechanism. Establishing such a unique number is one of the features of the Genealogy Registry System.

Integration

Integration is closely related to the uniqueness factor provided by the person ID number. If we identify each person by a number which is unique worldwide, then all data about that person can be linked together using that number, no matter where that data resides. Connecting all useful, legitimate, and relevant data to the single correct person is one way to cut down the chaos and confusion of records and people. With this process we can actually check off each public record as it is associated with a specific individual. When that process is finished, we can be reasonably confident that we have made good use of all the public records, and that they are logically connected with the correct people that they identify.

The Genealogy Registry System provides several mechanisms for establishing and preserving all the many types of links to the world's data which make this Internet-wide database system possible and valuable.

Fairness

An online system is needed to help everyone in the genealogy industry be as fair as possible in their dealings with other genealogists. If these fairness issues (which sometimes may involve data privacy and security issues) are resolved, then it should be possible to take full advantage of the new productivity made possible through cooperation. On the other hand, if people are not convinced that the system is going to be reasonably fair to them, they may be unwilling to cooperate, and the whole concept may never get off the ground, and everybody will lose.

How do we reconcile the two concepts of "genealogy should be free" and "the laborer is worthy of his hire?" Obviously, regardless of the oft-heard "free" sentiment, most genealogy is not free. Someone has to put in many hours to compile it, and may have to pay significant research expenses along the way. Those who support "free" online websites are usually paying out real money to keep those operations alive. It may be free to some people, but expensive to others.

It is one thing to be technically able to cooperate on an industrial scale using a mechanism such as I am offering. However, it is quite a different thing for people to actually want to use this cooperation capability. It would help if they felt that they and their contribution would be treated fairly. I keep hearing of cases where people have done a lot of genealogy research work, but they are not very willing to share it with others, for numerous reasons. I will list some of those reasons here, but I'm not sure that I know enough to list them all.

I assume that in most data exchange transactions, they are quite unbalanced. That is, one person gets the benefit of another person's work, without offering much more than a thank you. It is theoretically possible that, in the long run, it all balances out, and everyone receives as much help from other people as they offer to others. However, I think in practice that kind of balancing only rarely happens. Probably the more usual case is that someone has done a lot of good work and everyone else who hears about it wants to get a free copy, and these people have little to offer in return.

In many cases, the data was compiled by professionals paid by the original researcher, but the person requesting that data is unwilling to share any of the costs, and simply wants it for free. Even if the work was not done by paid professionals, the very extensive work by the original researcher, possibly including extensive research costs and travel costs, might make that original researcher hesitant to give away the product for free, especially if the data receiver is in a position to pay for part of those research costs but is unwilling to do his part.

Besides the unreimbursed value of the data itself, the work involved in doing the sharing may be quite significant. Answering a continual string of e-mails may require a great deal of work locating and extracting electronic copies of materials, or may mean sending expensive paper copies of some materials. Government and commercial organizations typically charge substantial fees for such information. The workload of answering e-mails may keep the researcher from extending his own research, so he may decide to greatly curtail his e-mail answers.

Some researchers have expressed irritation about how their data was used once it was transferred to another person. Perhaps those people receiving the data treated it as their own and did not give proper acknowledgment of the huge amount of work done by the data supplier. Or maybe they changed the data according to their ideas and placed it in some public location where the changes were a constant thorn in the side of the original data supplier. And, typically, that original data supplier would have no reliable way to correct or contest or annotate these online representations of their work.

Or perhaps the data receiver put the data in a online system where the very process of placing it there distorted or degraded it in some way, again leaving the original data supplier to fume, with no option for correcting it or at least stating his reservations as to accuracy, etc.

A nationwide genealogy project

All of these fairness issues become even more important when the goal is to finish the basic genealogy to a high quality standard for the entire United States. That is a huge job, but it can be finished very quickly if large numbers of people are willing to cooperate. However, if people see no value to themselves in having an entire nation's genealogy completed, or are concerned about not being properly recognized and rewarded for the work they do, then we may not get the needed cooperation and the project may be only a shadow of what it ought to be.

For each person to get a full 10-generation pedigree out of this project, that requires there to be about 1000 other people doing work that relates to his pedigree lines which he has not personally covered. So how many participants does it take to ensure that all participants receive the 1023 other surname lines they need? I believe that number is somewhere in the range of 30,000 people to 100,000 people, depending on how much time they are willing to put into their part of the project.

(For more localized projects such as a Utah community, where the numerous original pioneer families have been intermarrying for a century, and where almost everyone there is related to everyone else, a much smaller number of participants would be enough to make sure that everyone received all the data necessary for a full 10-generation pedigree. For example, if everyone involved shared 20% of their surname lines with everyone else in the localized project, then it would only take about $5 \times 1,024 = 5,120$ total participants to accomplish everyone's goals.)

Here are a few of the computational considerations for getting a complete team of people to finish the United States genealogy. The first assumption is that the number of historical people, those who have died in the United States up to a certain year, is about the same as the number of names recorded in a census for that year. For example, there have been about 210 million people who have died in the US up to 2010, and the current population is about 310 million.

From the 2000 census we learn that there are 151,671 surnames with at least 100 entries in the census, and that 1720 of those surnames represent about one half of the people in the United States. According to the 2000 census, the population was 281.4 million. The 2000 census surname data table accounts for 242,121,073 people out of the 281.4 million counted. Incidentally, the 2010 Census reported 308.7 million people in the United States, a 9.7 percent increase from the Census 2000 population of 281.4 million.

Our main goal is to complete the 70 million people who died before 1930. If we assume that since 70 million is about one fourth of the 281.4 million people counted in the 2000 census, perhaps we can also assume that the surnames were the same in 1930, except there were only one fourth as many people for each surname.

We might hope that there are about 150,000 people who are willing to collect all the data on their surname and add it to the central database. The only difficulty with that is that some people would be doing 100 names or less, while others, such as members of the Smith surname group, would have 2.3 million names to cover. Obviously they would need some help from their many cousins on that part of the project. If we assume there were only $(2.3 \text{ million} / 4 =) 0.57$ million Smiths up to 1930, that makes the Smith part of the project easier to complete.

If we looked at it a different way and said every person who participates should do 1000 names, whether as one part of the Smith contribution, or as a combination of 10 of the smallest surname groups, we would then need about $70,000,000/1000$ equals 70,000 participants. If we up the contribution to 5000 names per contributor, then we would only need $70,000,000/5000=$ 14,000 participants.

Individual benefits

So what are the benefits to an individual participant of working on a nationwide genealogy project? I should first mention that the LDS Church reports it has 130,000 people working as volunteers on its "Online Indexing" project to transcribe and index the 2.5 million rolls of microfilm in its granite vault, which contain 2.5 billion images and may represent up to 25 billion names. It also reports that it has had 7 million hits a day on its Family Search.org website. This doesn't tell us the reasons that people do this kind of work, but it does tell us that there are probably a large number of people interested in assisting genealogy work to be completed on a nationwide scale.

The individual participant can expect to get a complete 10-generation pedigree, which is 128 times as much data as he might have after completing the research on his pedigree going back three generations. If this participant does quality work, then there is a good opportunity to receive significant amounts in fees from other people who use his high-quality research work. Data and money are the direct individual rewards to be expected.

There are more generalized returns for those who take a slightly larger view of life. If a person would value having available a highly detailed personal history of everyone who ever lived in the United States, based on linking together all of the public records about that historical person, then it may be of value to that participant with the larger view. For example, one person's black ancestor who went from being a slave to being a successful businessman would be an interesting and inspiring story. If one could trace similar events in the lives of other people coming from the same time period, that might offer some additional dimensions to the story of one's direct ancestor.

Even more generalized and generic is the effect of having many people in the United States learn more about their ancestors and the struggles of life they passed through. Reviewing the hardships and triumphs of their ancestors might inspire them to live more righteous and productive lives, which in turn is a benefit to our entire nation. Simply having people today be less self-centered would be a benefit to all.

Going back to some financial computations, in the past the idea of doing genealogy research with the intent of supplying it to others was so costly as to be hard to imagine it as a common gift. However, with this new cooperative system, which can cost only 1/2000th of the past methods, it becomes feasible to make gifts of genealogy data to others. For example, if the cost of the results are in the range of \$1 per name, a person might reasonably consider offering a 4-generation pedigree (\$32) or a 5-generation pedigree (\$64) as a gift to family and friends. These figures are in the range of the cost of many books available at a bookstore, so they might be seriously considered as gifts to friends and relatives. LDS participants might find this gift-giving idea of particular interest, since many of them are genealogy enthusiasts, and they might use gifts of genealogy as a way to introduce their friends to some of the Church's activities. I am told this giving of genealogy as a gift is already a common event in some parts of California.

If we were to view the process of completing the basic genealogy for the entire United States as a series of gifts by all the participants to each other and to the nation, we might compute the cost of those gifts in this way: if each of 4 million genealogists spends two weeks of work on this project, and we assign a cost of \$10 an hour to that work, we might compute a cost of \$800 for each participant. Of course, the 4 million participants are expected to receive far more in data and in actual money than they put in. In fact, when

they have put in their unit of data, they should receive back 1000 times that amount of data, which they can then use to their individual benefit.

Extending the project worldwide

The great interest in genealogy in the United States, plus the personal resources which are available to invest in genealogy work, make the United States the natural place to start a worldwide system. Beyond the basic operating costs, there should be enough extra income from United States operations that it should be easy to spark the same kind of aggregation of high-quality genealogy data in Europe, and then gradually move to other places in the world as well. Since the United States has such a long history of immigrants coming to this country, the genealogy research collected in this country should extend to many countries in the world, especially Europe. Hopefully, many people who participate in this project will do it partly because of their idealistic and altruistic impulses to see a high-quality worldwide genealogy database constructed. Many of today's living people will not know all the places where their genealogy will lead them. By supporting this project, they can eventually find all their family ties to all the other countries of the world. Hopefully, that realization will cause them to be willing to make at least small contributions of money and data to the project.

The Genealogy Registry System promotes the maximum level of fairness in the genealogy industry by expanding and simplifying data exchange, and helps even further by providing a fee exchange system which should reward those who do the most valuable work with the most financial return, which should at least repay them for their research expenses, if not more. In an all-volunteer genealogy system, it is very difficult to get anywhere close to complete fairness, but adding the financial component fills in many of the fairness gaps. The system also provides for one person to add notes and data to the work of another person, without in any way changing the original entries of that other person. This at least allows researchers to be sure their differing views and interpretations can be preserved.

3. The Hidden Benefits and Cooperation Power of Descendent-Sequence Research

An important central concept in doing the most efficient genealogy research through cooperative methods is the need to do most research and recording work in descendent sequence and form. This will be counterintuitive for most people, so I will spend a few pages trying to describe this important element of the general concept.

First of all, there is good reason to believe it is from 30 to 100 times faster to extract names from public source records if it is done in descendent sequence. See computations below. For a serious participant in this new process, a typical operation would be to take the time to gain an expert's understanding of a particular set of records, and then proceed to extract all names that have a particular surname. For example, in my case, I would want to find every occurrence of the Huff surname in a set of records, and make sure that I documented and interrelated each of those names to the extent possible. This process would be especially suitable for large record collections such as a census. For smaller collections of records, it might make sense in some cases to simply transcribe and interrelate *all* the names of that particular record set. The point of this activity is to extract substantial units of data and use it and display it in such a way so that no one else ever has to perform that particular task.

In the past, we might have observed thousands of researchers each looking sequentially through a microfilm copy of a particular set of old records, with each one of those researchers intent on finding one or a few names of interest to them. From an overall standpoint, this is an extremely inefficient process. Wherever possible, it would be best for a person to convert a large segment of that set of records, or perhaps the entire set, into a form suitable for general use, so that no one else needs to repeat that laborious process.

I was interested to chat with one genealogist who was doing research in Germany. She was often looking for her people in small villages, and she had decided that whenever she found a village which contained one of her ancestors, she would extract all the data about the entire village. All of those people were certainly her relative's friends and neighbors, and many of them might be part of her family. She might later find that she had indeed included numerous relatives in the extra names she had extracted. The only shortcoming to this process was that she had no good way to share this bounty of data with other people so that any other people who needed to look at those villages would not have to repeat all the work that she had done. One purpose of the new computer system is to allow the genealogy community to get the most value out of this extra work which people do.

Efficiency estimates and comparisons

Descendent-sequence research appears to be from 30 up to 100 times faster at the point at which names are extracted from a public record set. I don't have extensive statistical data on this point, but from the anecdotal data I do have, I believe it to be correct. As an example, if someone can find 100 names in a particular record set using the new procedures, where they might have stopped after finding only one person using the old procedures, and the preparations for being ready to use that record set were the same in either case, then it might easily happen that they ended up collecting names 30 to 100 times faster.

I have one valuable illustration in the form of the book *Descendants of Engelbert Huff*. When I first became acquainted with the book, I was astonished that one man could have accomplished such a work. It is

obviously completely out of the question for a pedigree-sequence researcher to ever compile 15,000 names. But this man did it. I know that he had help from many other people, especially in the collecting later generations, since I was one of those who contributed family data to the book. But that kind of cooperation is also one of efficiencies which other people can duplicate in descendent-sequence research. The fact that he was focused on a single-surname descendent structure made it possible for him to seek the cooperation of many others, so that his personal time devoted per name could be kept in the one hour range. This example of great efficiency was part of what encouraged me to study the mathematics and practicalities of improving genealogy systems. Surely, if he could do such a fantastic job manually, and mostly by himself, it must be possible to do even better and faster work with specific computer support and even more cooperation.

Besides the book I just mentioned, I also interviewed a pedigree-sequence researcher about her experience. We can compare these two research examples to gain some insight into the difference between the efficiency of pedigree-sequence research and descendent-sequence research. In each case the researchers spent large amounts of research time covering about 10 years. At 1500 hours of work each year, that would be about 15,000 hours in all for each researcher. The pedigree-sequence researcher went back six generations on all lines and 10 generations on two lines. In that process she would have found about 144 direct ancestors, spending perhaps 104 hours per name. (15,000/144 equals 104 hours per name.)

The descendent-sequence researcher compiled about 15,000 Huff names in all, starting back 13 generations from the present and coming forward. That obviously allows him only about one hour per name for the work he did. That indicates that he was approximately 100 times as efficient as the very determined and accomplished pedigree-sequence researcher I interviewed. He did a very high-quality job and published a book, *A Huff Genealogy: Descendants of Engelbert Huff of Dutchess County, New York* by George Lockwood Trigg, Heritage Books, Inc., Bowie, Maryland, published in 1992.

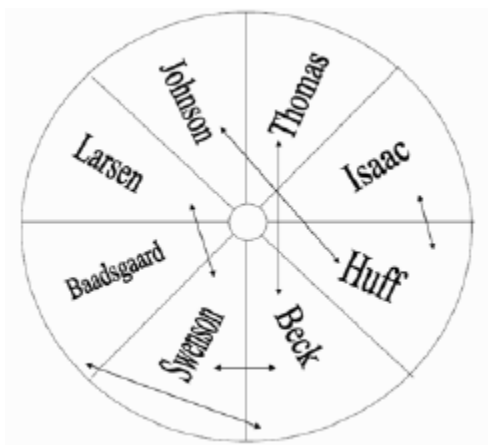
The end of duplication through descendent-sequence research

In the typical pedigree-sequence research process, there may be thousands of isolated people who each discover a particular name, and record that name with minimal connections to the rest of that person's family. These individuals are often seemingly presented as though the only people to whom they are related are their spouse and the researcher who found them, plus the thin line of one child in each generation (and that child's spouse) down to the present. In other words, the typical pedigree-sequence process creates potentially tens of thousands of duplicate entries for each of these unique historical people. We have many times seen the chaos which is created when computer programs are used to try to reduce the duplication by matching and merging these names and all related names. But this chaotic aspect of a database created by pedigree-sequence research is completely unnecessary. It is simply an artifact of the pedigree-sequence process.

In the descendent-sequence research process, there should never be any significant duplication created in the process. If one person, or a cooperating group of same-surname cousins, is putting together a descendent structure beginning with a common ancient ancestor, then each of the same-surname descendants documented will be clearly known to that single researcher or to the closely associated group of researchers. They will all certainly know whether there are any duplicates in the set of names they are collecting. There is very little chance of duplication to begin with, unless there are some unusual cases of suspected twins, or adoptions, or some other unusual situation. And since, by definition, this researcher or group of researchers are the only people in the world working on that set of descendants from that ancient ancestor, there cannot be any duplicates anywhere else in the world. If for some reason two or more groups of researchers do indeed begin to work on the descendants of a particular ancient ancestor, the

computer system should help them quickly become aware of those other groups. They should then quickly be able to work out an efficient cooperation plan, or, if nothing else, agree to stop entering competing information that can cause confusion. If all else fails as far as working out a sensible cooperation plan, the computer system has the option to simply make one of the sets of entries invisible to the public and end the duplication problem that way.

The internal format of a descendent-sequence database



This diagram of the "pizza pie" database, where data is entered primarily in descendent sequence, is a very important part of the system concept and appears in several places in this book. Each of the slices of this "pizza pie" diagram is a descendent structure, beginning with an ancient ancestor and coming forward in time showing all the same-surname descendents of that ancient ancestor. Notice that the women will typically appear twice in this database, once as a daughter and once as a wife. There are "same person" arrows on the database diagram which represent connecting these women as they appear in their birth family and in their marriage family.

The basic idea is that this descendent-sequence database can be constructed hundreds of times faster than any pedigree-sequence-based database, and then all possible pedigrees can be simply read out at the end.

This structure assumes a researcher focus on a single surname, usually the same surname as the researcher, and allows a semi-automatic way to divide up the nationwide research project which encourages people to specialize. Most people do have an interest in their same-surname ancestor, and will usually have interests concerning their same-surname cousins, especially if they have grown up in close proximity to them in a clan-like situation. As just explained above, this methodology also has the great advantage of essentially ending the introduction of duplicates into a database. The descendents of a particular ancient ancestor are a completely fixed and determined group, who lived and are documented in an extended family setting, and are not usually subject to all the difficult interpretations as to which person is which that pedigree-sequence researchers might encounter. Obviously, people and families live their lives in descendent-sequence, and public and private records about those lives are also recorded in descendent sequence, so it should make some intuitive sense to retrace those steps to most completely and accurately collect data about those people.

As I mentioned earlier, following a family in descendent-sequence ensures the researcher that all individuals will be found and documented. In contrast, those who follow pedigree-sequence research methods almost automatically exclude approximately 20% of all historical persons. That 20% are those who

had no progeny, or had no progeny which continued down to the present day, and so, by definition, will have no living descendent to find them as part of that living person's pedigree.

Other possible applications of the descendent-sequence database concept

1. Use less computers: A completely manual genealogy system

I find it interesting to note that the basic process which I describe could have been done 100 years ago using completely manual methods. By my computations, if we had 2000 file cabinets and 500 filing clerks in one big room, about the size of a genealogy convention vendor exhibit space, we could have completed the basic genealogy for United States with about 17 years of work. Obviously, computers can do that work much more quickly and possibly more cheaply today, but it would have been a highly successful project, producing high-quality data in a timely manner, if we had started long ago. Of course, no one realized that it could have been done back then, so now we can use the more capable and efficient computer methods to complete the task. One of the restrictions on the manual system would have been that the supporting source records could not be stored in close proximity to the basic identifying and relationship data without the manual filing system growing to a probably unsupportable size. Those kinds of size, capacity, and performance concerns are easily overcome with today's computers.

2. Use more computers: A supercomputer with no genealogist assistance required

One method mentioned on genealogy technical blogs assumes that there are too few genealogists, and even those few are too poorly trained, to ever complete the nation's basic genealogy research, and that only a properly programmed supercomputer can ever do an adequate job of assembling completed genealogy from existing public source documents.

My reaction to this idea is that as long as we expect human genealogists to be involved in genealogy research (and it is hard to imagine genealogy without genealogists), I believe the Genealogy Registry System concept I am describing in this book is the best that is available. The method I propose could also mean that the entire nation's genealogy is completed decades before this other more esoteric method could even be attempted.

I assume that this supercomputer method would operate much like the "descendent-structure workstation" feature that I describe in Chapter 5, but on a larger scale, and with no genealogist input along the way. The Genealogy Registry workstation feature, made logically possible by a descendent-sequence approach to genealogy research, uses computers to gather all potentially relevant data about a descendent structure for a specific surname, usually collecting data in the 1790 to 1930 time frame, and placing all of it on a researcher's PC, including transcriptions, indexes, and record images. With all of that data instantly available to the researcher, then the process can begin of analyzing the documents and piecing together one or more descendent structures from that data.

To continue, the supercomputer proposal would almost completely replace human genealogists with a giant computer which would begin with completely transcribed and indexed versions of all public genealogical records, and then stitch all of those public records together to construct an entire nation's genealogy. There is definite merit to this proposal, especially since the finished product would be completely based on source records, making it as accurate and reliable as current records would allow.

The great difficulty is the human element. At this point in our technology, we still require that human eyes and brains process the vast number of old public source records so that accurate transcriptions and indexing can be accomplished, preparing all this data for effective use by people and computers, present and future. And even after the supercomputers have finished their work of piecing all of these documents together, there would still remain the rather significant task of verifying that the results do in fact make

sense to humans. And there is also the rather basic question of who would be willing to fund this rather expensive and highly centralized project.

In the Genealogy Registry System I describe, we can have millions of people following their interests and hobbies to both help with transcribing and indexing the source documents, and evaluating the results, including writing a critical analysis of each piece of evidence that may be linked to a person. These higher level evaluations of research results, based on factors and intuition which are far too complicated for today's computers to deal with, are of great value in getting a high-quality, dependable result, and also just happen to be the sort of activity which many people enjoy doing. Many of these people would feel very skeptical of some machine doing it all for them. Would they really want to hire a room full of computer people to take credit for doing the nation's genealogy? Also, there is an exceedingly high possibility that much of the data produced by such a machine process would be only marginally useful. Other similar but lesser attempts have been quite disappointing. The machines can do no better than the quality of the data they receive allows. Humans have this nice ability to be able to upgrade the value of information they receive as part of the process.

We might consider the possibilities of whether this supercomputer method of completing genealogy might happen anytime soon. The LDS Church has a large program under way to convert perhaps 2.5 billion source record microfilmed images into Internet-accessible digitized form and then transcribe and index the contents. If I were to guess that 10% of the work has been done already in the last five years, that would mean it would take another 45 years to complete. I assume the pace in the future will be faster than that, but it might still take another 10 or 20 years to complete, especially since I hear that there are legal constraints on large segments of the data which keeps it from being made completely available online. When all that preparation work has been done, then the computer matching could begin.

Another alternative would be to use the databases of Ancestry.com. It is possible that those databases would already be complete enough to allow this supercomputer matching process to begin. However, it seems unlikely that Ancestry.com would be willing to do this expensive project itself, and, in the process, make a huge change in its business model which has been successful so far. If some organization wanted to gain access to those databases to run this supercomputer experiment, Ancestry.com might reasonably charge them a large sum, perhaps up to \$3 billion for that access. If Ancestry.com is receiving about \$300 million a year from its data subscribers, totaling \$3 billion in 10 years, and that flow of revenue might be cut significantly if all the nation's genealogy research work had already been done by a supercomputer, then it might be facing a significant change to its business or even the end of its business as it operates today.

In the meantime, the Genealogy Registry System project which I suggest could be started and completely finished, so it wins over the other project on several practical grounds.

4. Cottage Industry Versus Industrial Revolution Methods

It is important to point out that the traditional research methods of genealogy, which are still the prevailing methods today, are basically the cottage industry methods in use before the Industrial Revolution. Those cottage industry methods are characterized by isolated individuals carrying out all necessary steps to complete a product. There is little or no specialization, and each worker is required to have a vast amount of knowledge in order to be effective, usually requiring a long apprenticeship.

The most striking example I can think of would be using cottage industry methods to create an automobile, as opposed to using modern-day industrial methods. For the cottage industry methods, one might have someone dump piles of metal and plastic, and perhaps wood and fabric, in front of the workman's garage, and that craftsman would then be required to make every needed part from scratch from those raw materials. The complexity and sophistication of the final automobile would be greatly limited by the knowledge and experience of that particular individual. It might be that this person could never produce anything more complex or sophisticated than the 1897 Stanley steamer with its partial dependence on the preceding bicycle technology, and with a maximum speed of 25 miles an hour. It would be little more than a horseless buggy, with none of the amenities of a modern automobile. Certainly it would be impossible for one man to ever know enough, or live long enough, to produce a 2011 Lincoln Town Car as we know them, complete with air-conditioning, radio, CD player, GPS, etc. We can assume that it would take a very long time to create each car, perhaps many years of a craftsman's efforts, and the resulting product might be of low quality and of low complexity and low performance.

Compare this with a trip to a large Ford plant where one could specify the parameters of the car desired, including color, engine size, and many other features. After perhaps a two-hour wait, out would pop a beautiful, shiny, highly sophisticated, high-performance automobile, guaranteed to work perfectly for at least 100,000 miles of safe and comfortable travel at a speed of perhaps 75 miles an hour, but capable of speeds up to 150 miles per hour if needed.

Using modern industrial methods, the various components of an automobile are broken into perhaps 10,000 separate pieces, and each of those pieces is contracted out to a specialist who can produce very high quality products in an efficient manner. All of this takes an enormous amount of planning and coordination, but the results are spectacular. Each of these high quality items is gathered together in an assembly plant where they can be incorporated into a finished product over perhaps a period of two hours, as the assembly line inexorably moves from the beginning to the end.

There are those who have bemoaned the "dehumanizing" nature of repetitive factory work, but those who have earned the high wages of being part of such an efficient system, and enjoyed the low cost and high quality products, are not likely to be those who are doing any of the "bemoaning." The important feature which most people understand is that by making people hundreds and thousands of times more productive, everyone in the society wins by becoming more affluent.

People may complain about Walmart these days, which is one of the latest applications of industrial organization, but very few people are willing to pay two or three times the price at an inefficient store when they can get the lower price at the highly efficient store.

One might notice that the main people who complain about these industrial efficiency advancements, are the less efficient competitors who are being left behind. They may couch their rhetoric in the form of

protecting the good people of the nation from this rapacious new entrant into commerce, but the truth is they simply hate to lose out to a more efficient competitor.

Many people have heard of the Luddites who tried to resist the advancement of factories by physically attacking and destroying the machinery. They may indeed have been trying to preserve their now-obsolete cottage industry production methods, but by resisting improvements in efficiency, they were hurting everyone in the society. At nearly every step in improving efficiency of production and the capabilities of the product, there is someone who is missing out, or being bypassed, and who may react like the Luddites. But from today's vantage point, looking back over history, during which we have gone through many generations of industrial productivity improvements, we can see how primitive things would have remained if these Luddites had been able to stop industrial progress. The buggy whip makers were not happy to see the rise of the horseless carriage, and the end of their craft, but would the general populace want to let this small group of people limit industrial advancement for everyone else? We take these enormous efficiencies for granted in our industrial age, and perhaps forget the huge gap in productivity between cottage industry methods and industrial methods.

Wealth of Nations

In his famous book *Wealth of Nations*, Adam Smith estimated somewhere between a **240- and 4800-fold increase in individual and group productivity** by dividing the labor in his pin factory case study. It is this same kind of boosted productivity which we are looking for in the genealogy industry. Obviously, the detailed methods to achieve that productivity are different, but the general concept is the same.

Ford Motor Company

As another illustration, the historical statistics show Ford as a latecomer to the automobile industry, with its first production statistics showing 1708 automobiles constructed in 1903. But it quickly rose to produce nearly 10 times as many cars as the next competitor, as in 1921 when Ford produced 1,275,618 cars and Chevrolet produced 130,855 cars. Obviously, in an industry which was becoming increasingly more productive, Ford was still 10 times more productive than the next competitor. It is time to have the genealogy industry take lessons from its predecessors in the material world, and take advantage of these huge jumps in productivity which are possible now.

Oxford English Dictionary

The compilation of words of the original Oxford English Dictionary provides an interesting case study that relates somewhat to collecting genealogy data about historical people. The compilation process was begun in 1860 and the completed result was finally published in 1928 with "over 400,000 words and phrases in ten volumes." (The current OED edition contains approximately 600,000 words.) It was originally compiled by asking 800 volunteers to read a library of books and prepare entries and quotations for words they found. There was obviously an enormous potential for duplication in this process, since how could 800 readers each know what had already been submitted and compiled by all past and current readers and editors? There was certainly no online computer system to show a participant the exact status of the entire project at any time, as the Genealogy Registry System can for a nation's genealogy. That is presumably much of the reason why it took 68 years to finish the OED1 task. Finishing the nation's genealogy is a much bigger task, of course, since documenting 70 million people is far more complex than the 400,000 words of OED1, but we now have sophisticated computer mechanisms so that we can exactly coordinate the work of 4 million participants, which is 5000 times more people than participated in the OED project. Note that by using today's computer technology, the OED1 project could theoretically have been completed in 5 days. Computation: 68 years*365 days=24,820 days/5,000 times more people=4.964 days. For historical background, see http://en.wikipedia.org/wiki/Oxford_english_Dictionary

5. Other Important Features

Previously unrecognized opportunities to increase efficiency

It is certainly helpful to solve the recognized historical problems of the genealogy industry, but having done that, a previously unrecognized set of opportunities to increase efficiency and increase project value also presented themselves. Here are a few of the larger ones:

1. Get all world's genealogists involved in this naturally worldwide project for another huge boost in productivity and efficiency. Add up to 80 million worldwide genealogists to the perhaps 4 million genealogists in the US. That could be 20 times better than the US alone, and 800 times better than the perhaps 100,000 LDS genealogists.

I have focused my calculations on completing all the basic genealogy work for the United States, but all the opportunities and facilities for completing the United States apply at least as strongly to the rest of the world. As various genealogy mathematical explorations have determined, all of us are related to each other in ways that we might not suspect. For example,

The mathematical study of genealogy indicates that everyone in the world is descended from Nefertiti and Confucius, and everyone of European ancestry is descended from Muhammad and Charlemagne.

All Europeans alive today have among their [common] ancestors the same man or woman who lived around 1400.

20 percent of the adult Europeans alive in 1000 would turn out to be the ancestors of no one living today (that is, they had no children or all their descendants eventually died childless); each of the remaining 80 percent would turn out to be a direct ancestor of every European living today.

We're all descended from Charlemagne. But can you prove it? That's the game of genealogy. See: Steve Olson, "The Royal We," *The Atlantic Monthly*, May 2002, pp. 62-64.

It would be ideal to get all the genealogists in the world cooperating on the same project, since everyone in the world will benefit from the combined database they could create so efficiently.

2. The "descendent-structure workstation" feature can help assemble raw genealogy data into descendent structures up to 200 times faster than traditional methods.

We now have the technical capability of quickly and automatically assembling on one user PC all the transcriptions and indexes of all public source records, plus all the actual source image records that are now in existence. For example, for the Huff surname, for which there were about 30,000 historical individuals documented in the 1930 census, it would be possible to start with the 1790 census and assemble all the census record entries up to 1930. This process could then go on to collect, on the user's PC, all known birth records, death records, marriage records, land records, etc., that deal with the Huff surname. This assembly process could go on at night or on weekends so that the user need not be involved in the thousands of detailed Internet operations. With that complete collection of all available data in one place, the user then could go on to examine multiple documents at once, operating at his own maximum "brain bandwidth" rather than be constrained by the difficulties of Internet operation, seeking out one document at a time

from various websites. Many PCs operate up to 400 times faster internally than the Internet can deliver external data to the user, so it makes sense to make that internal speed available to the researcher.

This feature is obviously dependent on the use of descendent-sequence research logic, which is the basis for much of this project. This is just another example of the mathematical power of using the descendent method. Obviously, when one is doing pedigree-sequence research, this kind of workstation feature is of little use. Where it is perfectly possible to assemble a full 12 generations of data in the descendent sequence, using a single surname, that is essentially impossible to do in pedigree sequence research. By definition, when one begins pedigree-sequence research, no one knows more than a few of the potentially 4096 surnames that need to be researched to achieve a full 12-generation pedigree.

Storage features

From the user's standpoint, it is highly desirable to have all source documents available through one integrated portal, which handles all addressing issues. The data storage plan of the Genealogy Registry System is to link to and use publicly available source record images wherever available, as the LDS Church plans to provide, as a means of minimizing costs and duplication for the GenReg.com site. The GenReg.com site would also link to and use family websites where available, especially for large volume items such as videos, rather than try to store such voluminous items centrally. The GenReg.com site would emphasize offering a place to store all materials which are not satisfactorily stored elsewhere in, and some mild enforcement mechanisms would be employed to minimize storage redundancy.

The GenReg.com site will have special features available to minimize storage requirements. One example might be where 20 different people point to a single census record image, or a ship's manifest, etc., or any other documents that do not appear in an addressable form on the Internet. Those images might be uploaded 20 times in conjunction with individuals' research, but eventually there should be only one image, with 20 pointers to it.

Plans and features for publishing the results

For each historical person, the option is provided, and the participants are encouraged, to add a carefully reasoned synopsis and analysis of all available evidence, which may be several pages long. This document allows the serious genealogist to make sense out of the data which is available for the benefit of anyone else reading about those people. This document could be used as direct input to a book publishing operation. This material could be edited and reduced in size to fit the constraints of a physical publication, or retained at full-size for an electronic extract of the comprehensive database. For example, this set of document from family entries could go straight to a giant PDF, perhaps 1 or 2 GB in size, for distribution to family members.

Genealogy problems unique to the LDS church

1. The LDS Church needs a continual flow of fully-researched names for temple purposes, but has never had a system that completely meets that need. The Genealogy Registry System would allow the Church to completely fill that need for the foreseeable future.
2. The Church needs the ability to answer ethnic groups who object to the LDS Church using, in their temples, names in common use among those ethnic groups. One recurring example is Jewish names, especially those associated with the Holocaust, where Jewish ethnic defense groups have objected to what they consider Mormon interference with the Jewishness of the Holocaust victims who died because of their religion. One simple answer to these objections would be to show a family connection, from any particular name to be sent to a temple, to a living member of the LDS Church who might plausibly have the authority to give permission for those ordinances to occur. It requires a computer system with specialized features to solve that problem, and the Genealogy Registry System includes those features.

Making better use of our online source record resources

This new system offers the possibility for large blocks of genealogy data to be completed to a high quality standard by genealogy specialists, so that the work for whole nations can be done in a remarkably short period. At that point much of the work in the whole industry could shift from the basic clerical function of identifying individuals and their relationships, and move on to the far more interesting and fulfilling examination of all the "extra" history that might be available about their ancestors, as found in all the other records which might be available, including newspapers, etc. We might find that, on average, there are at least 20 entries in public records for each historical individual. Having once completed the basic matrix of people and relationships, researchers would then have the leisure to fill in all of these other sources of information.

All of this might have an interesting effect on the main data suppliers such as Ancestry.com. It seems likely that researchers today typically look only for the minimal amount of data which they think will substantiate the identities and relationships they are looking for. But with all the basic matrix of ancestors already constructed, they would then have reason to go back and make use of all the other records which these online data suppliers might have. Where they might have been satisfied with one record in the past, they might now have the time to seek out and analyze the 20 or 30 records which might be available from all sources for each of their ancestors.

The end of most "brick walls" in genealogy research

If people are willing to participate in this cooperative project as it is outlined, then that should mean the end of most so-called "brick walls" in genealogy research. If we quickly and carefully document hundreds of millions of real people, and associate with them the available source records, that alone will identify most people and remove the reasons for "brick walls" that pedigree-sequence researchers talk about. Again, the process is changed from the "lone wolf" researchers continually jumping back in time to find a name or two in widely scattered record sets that they usually know little about, to a highly cooperative effort where whole record sets are quickly turned into part of a large pedigree-linked database. The overwhelming bulk of historical people will be swept up in this process and linked to most of their relatives. There may be a few people who are missed in the process, or whose connections to any particular researcher are not clear, but when such a "missing" or "questionable" status is assigned to a person, there may be hundreds of people who want to solve that problem, not just one researcher, so it is highly likely that every such "missing" person will be found. This is like the typical thousand-piece puzzle process, where all the easiest pieces are linked together at the beginning – the edges, the corners, the most striking aspects of the picture itself, etc. – and then the harder, more obscure pieces are positioned last, when the number of places they might fit have been greatly restricted by the prior work. Or, to explain the process differently, each historical person will be looked at from multiple directions by multiple researchers using multiple sources, and it is unlikely that very many historical people will escape this scrutiny.

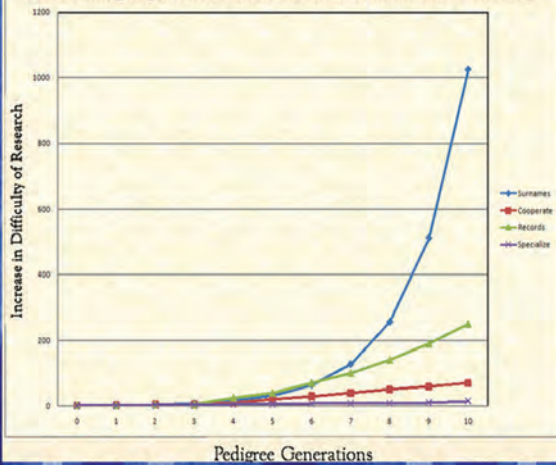
This might be a good place to mention again the fact that descendent-sequence research should find every person who was born and recorded, as opposed to pedigree-sequence research which usually misses the perhaps 20% of the people born into the world who do not, in turn, have posterity reaching down to the living today. These "people without posterity" are not going to be any of the "missing people" or "brick walls" that pedigree-sequence researchers are concerned about, because they are outside the set of people that pedigree-sequence researchers are seeking. I only mention it here to emphasize that only the descendent-sequence methods will catch everyone for whom there is an historical record.



Join the Genealogy Productivity Revolution



Overcoming Exponential Increases in Workload and Difficulty



Genealogy Registry

Old and New Paradigms of Genealogy Systems

Old: Isolated Users, seeking Raw (non-lineage-linked) data from central sites

New: Networked, Collaborating Users, sharing Finished data through the central site



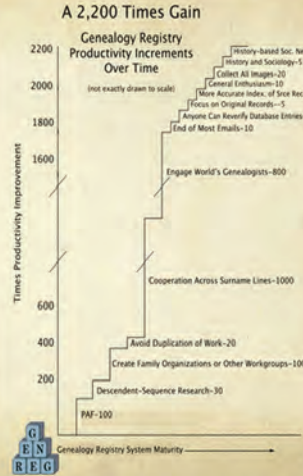
Genealogy Registry

Improve Productivity by 2000 times through cooperation

Cooperation is enhanced in 14 major ways

Benefits increase exponentially as the number of participants grows

www.GenReg.com



How Much Cooperation?

Each 10-generation pedigree needs data on 1024 surname lines.

If 1,023 other families complete their descendent research, your pedigree will be complete.

If 100,000 families participate, all U.S. surnames will be complete



The Descendent Method of collecting genealogical data makes possible a very compact and efficient database, with no duplication. It also allows the maximum cooperation among researchers.



Genealogy Registry



Do Genealogy



Through Cooperation



The Real Future of Genealogy

Finish U.S.

Genealogy in 2 weeks

Through Cooperation

(Industrial strength, 2000 times faster)

www.GenReg.com

The Genealogy Registry System

The project goal is to *finish* all basic U.S. genealogy within a two-year period. It can be done with the equivalent of 2 weeks work by all genealogists through vigorous, computer-assisted cooperation.

Calculation:

300 million people have died in the U.S. There are now 4 million active genealogists. That means each genealogist must enter 75 names (from their same-surname ancestors).

If we allow 1 hour to locate and enter each name, that adds up to about 80 hours in total, or about 2 weeks work for everyone.

Graph: A 2,000 Times Gain in Productivity is Possible Through Cooperation

Of the 14 efficiency elements shown here, the largest single element is entitled "Cooperation Across Surname Lines," which can render a 1000 times productivity gain. Participants enter one unit of data in descendent sequence and receive back 1000 similar units for use in building their full 10-generation pedigree (with 1024 surnames).

Database Format

All descendent structures are placed logically side-by-side. Connections are made between the places where women appear as daughters and where they appear as wives. Once all the data is entered and connected very efficiently, in descendent sequence, all desired pedigrees can be computed.

Kent Huff
huffkw@juno.com
home/office: 801-798-8441
cell: 801-615-9032
Genealogy Registry
1748 West 900 South
Spanish Fork, Utah 84660
www.GenReg.com (beta testing)
computer consultant, attorney, author

The Genealogy Productivity Revolution

All researchers are encouraged to join in the genealogy productivity revolution, where vigorous cooperation will bring the maximum advantage to themselves and all other participants. For the first time, cooperation is in everyone's self-interest.

Graph: Exponential Genealogy Difficulties and Their Solutions

1. The number of one's direct ancestors doubles at each generation going backwards, so that at generation 10 there are 1024 people, each with different surnames.
2. Using public records generally gets more difficult as we go back in time. The older records tend to contain less data, to be less structured, and harder to read. The researchers may also need to learn a new language to be able to do a professional job.

The solution in each case is to encourage researchers to specialize and cooperate so that these exponential difficulty curves can be flattened out for individual participants.

Old and New Paradigms of Genealogy Systems

Old: individual researchers are isolated from each other, and are dependent on central sites, where they mostly seek raw, non-lineage-linked data, because the lineage-linked data they find there is of low quality.

New: Researchers are able to cooperate on a grand scale and share large amounts of high quality finished lineage-linked data.