

The Genealogy Mathematics of Cooperation

How To Finish the United States In Two Weeks Using Cooperation and New Technology to Assemble Families 1000 Times Faster

ProgenyLink.com

The goal of this short article is to give readers a completely new way to look at the process of doing genealogy research. This is how Henry Ford would do genealogy, using specialization and cooperation to gain the huge efficiencies that come along with industrial cooperation processes. I hope to convince some genealogists to change their procedures a little bit and join in a cooperative project. I use the term "industrial strength cooperation" to describe the new process.

Most people who do genealogy research probably assume it is an infinite task, which can never be completed, and so we will have to wait until the millennium to finish any significant part of it. But I believe we could do all the names that we have records for now, if we would just get organized. What I hope is that by making the whole process much easier and faster, 1000 times faster, there will be many more people who will want to get involved and add their labor to the process, so that the entire project will continue to accelerate until completed.

Finish the United States in two weeks.

70 million people died in the US before 1930.

4 million genealogists times 18 names each = 70 million names.

Allow four hours per name = about 80 hours = 2 weeks work.

Other Options:

300,000 Online Indexing genealogists do 233 names each (in 6 months) = 70 million names.

5000 skilled genealogists do 14,000 names (in three years) = 70 million names.

The main goal of this project has been is to find a way to finish the basic genealogy for whole countries quickly. As I will show you in a minute, if we were all completely organized and had the right tools, we could finish the entire United States with about two weeks of work from each participant. I'm sure that sounds completely crazy to most people, but I hope that when I am finished you will see that it is perfectly possible, and I hope you will think seriously about getting involved to add in your part of the puzzle. As you can guess, it would take a lot more than two weeks to get all the genealogists of the United States trained on a certain method and have them agree to use it. But if they finally understand the project, then the results could happen very quickly.

I have spoken with a few others who have noticed that there is a very large amount of waste in what we do, and have sensed that we should be able to do it much better. However, even though several large genealogy organizations have spent hundreds of millions of dollars on computer software and databases, no one seems to have noticed the basic structures of genealogy mathematics which make it possible to achieve a 1000 times increase in productivity through a new form of cooperation.

I am sure almost everyone knows that there is a lot of duplication in genealogy research, and we need to do something about it. However, I don't think most people know how bad the duplication is now or how bad it can continue to be in the future. The really annoying problem is that, because of the exponential mathematics involved, the harder you work, the more duplication there is, which tends to cancel out the fact that you are working harder. This leads to a very frustrating situation where it's hard to make any noticeable progress.

Here is a calculation which shows how bad the situation could be if we try even harder to make progress: Let's pretend that all 300 million people in the United States suddenly decided they wanted to do 12 generations of their genealogy. That means each person in the United States would have to find 8196 of their ancestors. If they all finished their work, they would have found about 2.6 trillion names. But we know that there are only about 70 million people who died in the United States before 1930. So if we divide 2.6 trillion by 70 million, we get 37,000 times as the duplication factor. In other words, if we did our work individually, as we do now, and we all did 12 generations, on all possible lines, and put it all in one database, the names in that database would be duplicated on average 37,000 times each. That should give you a hint about why it is impossible to finish any large amount of genealogy using our present methods. Those current methods are just too inefficient and wasteful. The point here is that just working harder in traditional ways will never let us be successful. We have to change the procedure if we want to make any serious progress.

The Possibility for Duplication If There Is No Coordination Among Researchers

300 million people times 8192 ancestors (12 generations) equals 2.6 trillion names.

2.6 trillion names divided by 70 million pre-1930 people = 37,000 average duplication rate.

Cooperation Versus Isolation in Genealogy Research

The 10 Generation Project

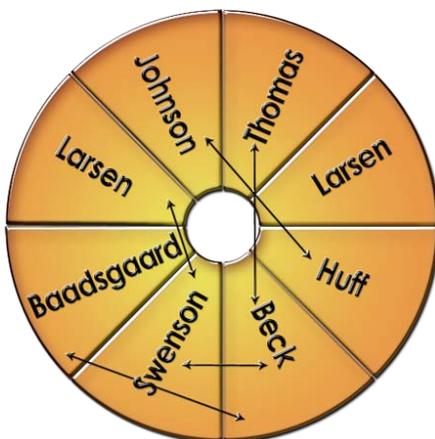
	Generations back	Surnames	Total names	Times more work to reach 10 gen	Lifetimes to complete (at traditional speeds)
	10	1024	2047	1	32
	9	512	1023	2	16
	8	256	511	4	8
Industrial strength cooperation is required to progress further. Assemble high quality names 100s of times faster.	7	128	255	8	4
	6	64	127	16	2
Possible to finish alone in one lifetime -- just a few generations.	5	32	63	32	1
	4	16	31	64	.5
	3	8	15	128	
	2	4	7	256	
	1	2	3	512	
	0	1	1	1024	

ProgenyLink.com - Action
ProgenySociety.org - Training
(formerly GenealogyRegistry.com or
GenReg.com)

Kent Huff
huffkw@juno.com, 801-798-8441

I have spoken with several people who are very serious genealogists. I tried to see how much work they were able to do with their time. One woman who lives in St. George has completed a full 6 generations of her ancestors and has taken two of the lines back 10 generations. I think she spent about 10 years of hard work doing that research. So she now has about 150 names of her ancestors. I believe she is a really exceptional researcher. **Most people seem to consider 5 full generations as a lifetime of work using current methods.** (See cut-off line on chart above). However, my experiments show that doing a full 10 generations, which is 32 times as much data as 5 generations, can be done more quickly with the new procedure than the five generations can be done today.

I want to contrast that lady's work with the gentleman who did the research for the Huff family. He spent a similar amount of time, about 10 years, and he produced about 5000 names of Huff-surname descendents of Engelbert Huff who was born in 1637. That is 13 generations of descendents. The book actually contains about 15,000 people, but the important piece is the 5000 people who have the Huff surname. If we do a little arithmetic we would see that 5000 names divided by 150 names = 33.33. In other words, he was able to collect high-quality names from original records about 30 times faster using descendent sequence research than the woman who was doing strictly pedigree sequence research. He also received some help from many of the Huff cousins, such as myself. That level of cooperation is hard to arrange doing only pedigree research.



Of course, he did not supply a complete pedigree for everyone in his book who is a Huff descendent alive today. He only did the Huff pedigree line for them. However he did it very quickly and very well – he has pictures and stories about a large number of these people. The goal is to take this highly efficient process, and make it so that, in the end, through cooperation, every participant gets a complete 10-generation pedigree. And how do they do that?

This next diagram (showing 8 surname lines) helps illustrate that if you connect the work of 1024 people, doing 1024 surname lines, then you can have a full 10-generation pedigree for yourself. That means you put in 1 unit of work and get back up to 1023 similar units of work for free. That could be a pretty significant payoff for

participating. Notice that each participant puts in his descendent structure of names, starting with an ancient ancestor. And then all of these slices of surname groups are linked together by the marriages that happened between those groups. Suddenly, the database is complete, and all possible pedigrees can be read out.

I'm sure some of you will be thinking "What if I do a good job on my names, with sources and so on, and the other people are lazy and do a bad job and don't supply any sources?" Well, we have some control over that in the system. If people put in poor data, we give it a poor quality rating, and some privileges are withheld from that participant until they improve the quality. This should encourage them to do a good job. If it is really low quality, we don't even make it visible to other people, until they fix it, so that we can avoid unnecessary confusion.

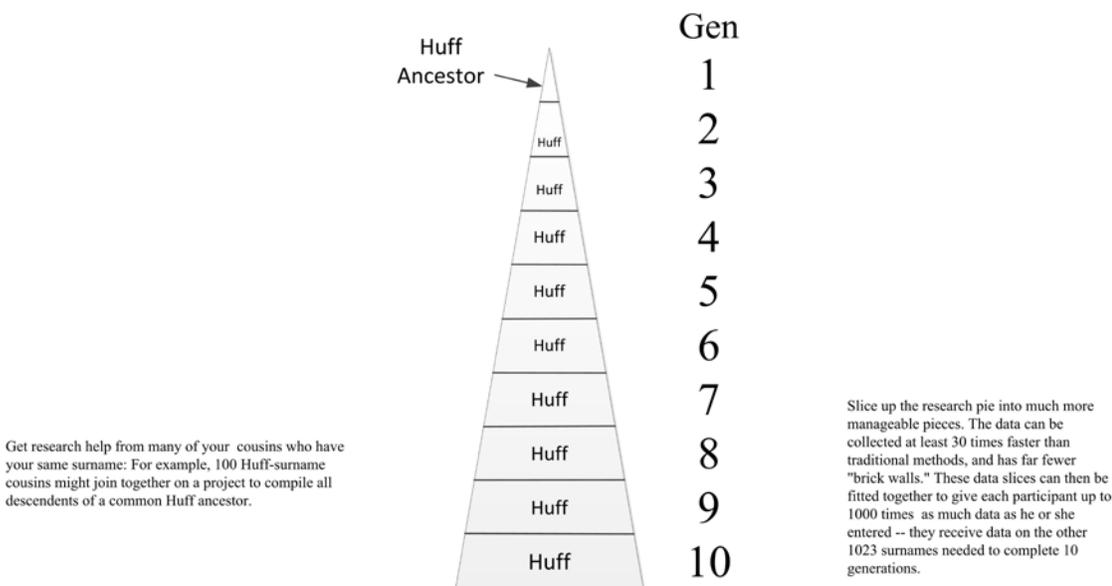
They could get to see your data, but not necessarily supply you good data in return, but we have a plan for that as well. If they use your data, they have to pay for it, either in points or real dollars. If you want to use their data because it is good quality, then you pay them. The idea is that we can keep it fairer this way through a fee-exchange process. (Obviously, if sources are available, then any researcher can quickly re-verify the accuracy and value of the data.) If people pay and get paid for good data, then it could all wash out to a near-zero balance for each participant. But if they only put in poor data which nobody trusts, then they get paid nothing, and they have to pay others for any data they get. That should give them some incentive to put in good data so that they can get other people's data for free. This is a little bit complicated, but we hope it will make things much more fair. Incidentally, if there is any revenue left over from this fee-exchange process, it will be used to pay people to fill holes in the database and to extend it into new countries.

We also make it so that you can have complete control over who can see and update your data. It could be only yourself, or just the people you specifically name, or everyone could see it, if you wish, or everyone can see it if they pay you to see it. This is a controlled cooperative system which lets you decide who can see the data. You can let people update it provisionally or immediately. A provisional update is one where people can enter data, but it does not take final effect until you agree. The immediate update, of course, takes effect immediately, but you can always go back and review any of the updates by any specific person in any specific time period, and decide whether you want those changes to remain or not. In other words, you can have complete control of your data, and yet you can get many people to directly help you load up your part of the database, without allowing them to do anything you don't authorize.

Here is what your portion of the database might look like:

Descendent Sequence Genealogy Research

Allows powerful cooperation without duplication



Participant's responsibility and contribution is much different and far more efficient when descendent research is used.

ProgenyLink.com
 Kent Huff, huffkw@juno.com,
 H/O 801-798-8441, Cell 801-615-9032

Everybody wants to know their complete pedigree. I know that someone is going to ask me how you can get the pedigree you want by doing just descendency research, even if it is more efficient in many ways? Well, that's the magic trick we use here. The whole idea is that you can put in high-quality names hundreds of times faster if you do it in descendent sequence, and when you're finished getting the names in properly, you can then read out all possible pedigrees. Obviously, someone will need to understand the system and have faith in it or they will never want to contribute to it. So my job, of course, is to show you how it really works. But here is a simple way to look at it: If you can get names into a large database hundreds of times faster by doing it in descendent sequence, and, in the end, get a pedigree which is at least 30 times as complete as anything you could do yourself in an entire lifetime, then it might be worth the trouble to participate. In other words, you can get your pedigree far faster by doing descendent research than you ever could doing straight pedigree research, even though the path seems more circuitous. I'm sure that sounds counterintuitive, but that's the way the system is designed to work. In fact, this kind of mathematics and indirect action is at the heart of all highly efficient mass production processes.

Future Shock – disruptive technology in science fiction space travel

Many years ago I read a science fiction story that seems to apply here. In this story the Earth built this giant spaceship which they were going to send off to go to the nearest star and other places. The idea was that it would be gone for hundreds of years and probably never come back to the earth. The spaceship had perhaps 1000 people on the ship, and it was expected that the ship would continue on for generations with one generation training the next to take over the controls and science experiments, etc., while sending any new scientific information back to the earth. In this story this spaceship takes off assuming it will never come back to earth and is gone for like 200 years on its journey.

Then one day a much smaller spaceship shows up right next to it. This smaller spaceship had left Earth about two days before and had caught up with the very large spaceship. This new Faster Than Light space flight technology offered the possibility that everyone could go back to earth if they wanted to, since other means were now available to explore the universe. This new space flight system had been made possible by the voyages of the large spaceship and the things they had learned.

But this new technology situation presented quite a social problem, because all the people on the big ship had been there all of their lives and could not imagine anything else. I forget how the story turns out, but I imagine that many people would be very sad to leave the only home they had ever known to go back to the earth.

Future Shock – disruptive technology in the genealogy industry

I assume we will have similar problems with the new genealogy research and cooperation possibilities. Many people may want to continue to do work they way they always have, simply because that is familiar, and they have complete control of what they do and the results they compile. Even though they could complete and share in high-quality genealogy hundreds of times faster using the new ways, it may take a while to decide that is what they actually want to do. I would not want anyone to think that I'm saying anything bad about past ways of doing things, because a great deal has been accomplished. But if we actually want to forge ahead and finish whole nations and finish our own genealogy back 10 generations, then we will need to do it a different way.

Professional genealogists might worry that this new system will hurt their businesses. Actually, I don't think it will, because there is still a very large amount of work to be done. But I do expect that they will have to change their procedures and perhaps change their business model to stay current. I could say a lot more about that, but I will save that for later.

Ancestry.com might worry that if people finish large amounts of genealogy, their \$400 million in annual subscription fees might be diminished. But I think there's a good answer for that worry as well. With current methods, we are happy to find one public document that provides useful information about our ancestors. With the new system, we could go on and find every public document that relates to our ancestors and link those ancestors to those documents. In other words, there might be 10 or 20 times as much use of the Ancestry database than we make of it now. This might cause Ancestry to change its programs and perhaps its business model a little bit, but it should still continue to be a very valuable resource.

Meeting LDS Church needs

Some of you might notice another benefit here. The Church needs perhaps 5 or 10 million names a year to keep the temples operating. Since most Church members do not do enough research to supply the temples with all the fully researched names needed, then the Church has to use substitute names which are not fully researched, meaning they are not linked to all of their immediate family members. If this new system were operating properly, the Church should have a nearly unlimited supply of completely researched names so that they would not have to take any shortcuts in data quality in supplying names to the temples. Also, if the Church decided to help with training and coordinating, then this cooperation process I describe could happen very quickly.

ProgenyLink.com

Kent Huff, huffkw@juno.com,

H/O 801-798-8441, Cell 801-615-9032