"Go forward and not backward"

Dramatically accelerating the Rate of Innovation in Family History Technology – Lessons Learned from Genomics and the Space Race

Luke Hutchison[1]

Abstract

Information technology provides an incredible raw toolset for accelerating the data processing pipelines of many data-intensive fields, including the basic sciences, but also holds great promise for the enabling of previously impossible progress in family history research. Family history technologies have not progressed as fast as many areas of information technology, but lessons on how to optimally accelerate innovation can be learned from the history of many different areas of technology. The RootsTech conference has been created in part to add critical mass to the development of family history technologies, however RootsTech is not purely a developers' conference, and increasing the rate of innovation family history technology will require concerted collaborative effort and cohesion within the diverse RootsTech community. Within the developer community, to avoid re-inventing wheels and to catalyze innovation and incremental progress, there is a great need for a centralized one-stop-shop for all things related to the family history technology development, incorporating a knowledgebase of problems, previous work and solutions, as well as mailing lists and shared source code repositories. Furthermore, in spite of the intrinsic need for cooperation in innovation, and in spite of the low inherent profits in the family history market, this paper also makes the case for *competitive* incentivization of innovation in family history technology, including through the creation of grand challenge prize competitions and software bounties.

[1]CSAIL, MIT. luke.hutch@mit.edu

Introduction

I completed a Master's degree in computer science at BYU between 2001 and 2004, where I undertook research in handwriting recognition for genealogical records with Dr. Tom Sederberg, as well as population genetics research for the Sorenson Molecular Genealogy Foundation[2] with Dr. Scott Woodward. My Master's thesis research was in tabular document image registration with Dr. Bill Barrett. While working in the BYU CS Department's Vision Lab, I was impressed at both the quality of the BYU CS classes and the caliber of faculty and students. I was also impressed with the large number of research projects that improved the state of the art in processing and analyzing genealogical data in many forms.

My peers at the BYU CS Dept had produced some incredibly competent Family History Technology (FHT) projects comprising thousands of lines of code requiring years of cumulative work. These FHT projects solved or partially solved a variety of important family history problems, including the ontological modeling of genealogical record structure, inference of family relationships with missing data, tabular document structure detection, microfilm ribbon scanning with automatic frame segmentation, fast search through microfilm images for handwritten word shapes using a flexible image matching algorithm, and many different image threshold, segmentation, contrast enhancement and noise reduction algorithms. Additionally, at the annual Family History Technology Workshop (FHTW), hosted at BYU, each year a number of interesting approaches to solving other genealogical problems were presented, such as name normalization, geolocation, record linking and merging.

After graduating from BYU in 2004, I lost track of what was happening at FHTW, but was asked to review a total of about 15 papers for FHTW between 2009 and 2010. I had the opportunity through reviewing those papers to observe what had happened in five years of FHT research. Unfortunately I discovered that—based on the papers I reviewed, at least—very little if any incremental progress had been made in that time. In fact between half and three quarters of the papers I reviewed were asking *exactly the same questions* that were being asked by my peers when I arrived at BYU eight years earlier in 2001—and yet there wasn't even a *single reference* to any of the research that had been done by anyone during my time there. In some particularly poignant cases, I knew that specific named problems were already *solved problems*, yet the authors were reinventing the wheel or dismissing them as too hard.

What happened to the thousands of lines of source code written by my talented peers in the early 2000s? Why did nobody know about the existence of the code, or even reference their papers? Why were these efforts still starting from scratch every time in 2010? The real problem is that there has been no consistent centralized hub or community framework for coordinating common effort towards solving these problems. FHT research will progress significantly faster with increased community cohesion and centralization of effort in solving the difficult technological grand challenge problems facing family history research.

With the creation of the RootsTech conference[3], we have a great opportunity to work towards that centralization of effort. At the same time however, with the merging of communities of users and developers into one conference, the RootsTech audience is both growing and diversifying. This will present unique and interesting challenges to building a cohesive community, but also presents opportunities for closer

[2] http://smgf.org/

[3] http://rootstech.familysearch.org/

interaction between the user and developer communities, which should result in software that is more useful and powerful to the end user.

Joseph Smith, Prophet-founder of the Church of Jesus Christ of Latter-day Saints (whose members are highly represented in the ranks of family history researchers), stated on the topic of seeking out our ancestors: *"shall we not go on in so great a cause? **Go forward and not backward. Courage...and on, on to the victory!**"* [4] This is an appropriate anthem for the FHT community: **Forward and not backward!**

Conquering the Challenges of Clan and Culture: The 10 'C's of Progress in Family History Technology

This paper discusses how to maximize the opportunities presented by the creation of the RootsTech conference. The first few points below discuss the interactions between the so-far disjoint communities of technology developers and users. The remaining points address encouraging and empowering the developer community to reinvent fewer wheels and rather to be making constant incremental progress.

#1. Cohesion (Community-building)

The Family History Department of the Church of Jesus Christ of Latter-day Saints has taken a big step in creating RootsTech, with the intent that it become the foremost conference in family history technology. The Family History Technology Workshop (FHTW) and the Conference on Computerized Family History and Genealogy have put their support behind and merged with RootsTech, and the FamilySearch Developers Conference has also been merged in. This has the potential to create an unprecedented critical mass of interest in family history technologies, but also brings together a very diverse audience.

The RootsTech audience consists of at least professors, undergrads and grad students from the BYU CS and genealogy departments, genealogy buffs, data normalization buffs, individual at-home hobbyist

---

[4]Joseph Smith continued: *"Let your hearts rejoice, and be exceedingly glad. Let the earth break forth into singing. Let the dead speak forth anthems of eternal praise to the King Immanuel, who hath ordained, before the world was, **that which would enable us to redeem them out of their prison*;** for the prisoners shall go free."—Doctrine & Covenants 128:22 (emphasis added).*

***** On the topic of *"that which would enable us to redeem them out of their prison"*, President Howard W. Hunter, fourteenth President of the Church of Jesus Christ of Latter-day Saints stated just days before he passed away, *"In recent years we have begun using information technology to hasten the sacred work of providing ordinances for the deceased. **The role of technology in this work has been accelerated by the Lord himself, who has had a guiding hand in its development and will continue to do so. However, we stand only on the threshold of what we can do with these tools.** I feel that our most enthusiastic projections can capture only a tiny glimpse of how these tools can help us—and of the eternal consequences of these efforts."—Howard W. Hunter, "We Have a Work to Do," Ensign, Mar. 1995, 64 (emphasis added). (Transcript of a fireside Honoring President Howard W. Hunter and the Genealogical Society of Utah. Pres. Hunter passed away on 03 March 1995.)*

programmers, representatives from small genealogy companies, representatives from large genealogy companies, framework/API implementers, framework consumers, webmasters, online genealogy forum participants, family history library patrons and staff and family history service missionaries. Some participants need to turn a profit in the family history market to keep their company afloat, while other participants are focused on doing *pro bono* work for its own sake. Participants comprise members, ecclesiastical and secular representatives from the Church of Jesus Christ of Latter-day Saints as well as those of other faiths. How can each participant of such a diverse community contribute to as well as derive maximum value from this community? It will be important for each stake-holder and interested party to work towards the creation of a strong, vibrant, friendly and cohesive community.

The interaction between the user and developer communities at RootsTech could prove very fruitful. When I worked for the Sorenson Molecular Genealogy Foundation (SMGF) after graduating from BYU, I created a tool to rapidly and optimally lay-out and render enormous sparse pedigree charts, with animated mouse pan/zoom functions as well as the ability to collapse/expand entire subtrees. I presented my work to SMGF's genealogists and quickly discovered that none of them liked or had much intention of using my whiz-bang bells-and-whistles user interface. Genealogists, it turns out, think of records as family groups, and these genealogists wanted to see a "standard" PAF-style 3-4 generation scrolling window view over a pedigree that allowed them to navigate one generation at a time by clicking arrows on specific lines. Personally I found the PAF user interface paradigm extremely confusing because I would lose my place due to the lack of a "big picture" context—but that's what they, the genealogical experts who actually had to use the program, were used to, and it's what made most sense to them. My opinion about what "made more sense" or what was a more optimal user experience was, in fact, irrelevant. I wasn't even an intended end-user of my own code! This is not an uncommon occurrence—in user interface design, developers frequently create a design that works the way they want it to work, not the way users want it to work.

On the flip side, users stand to gain a lot by allowing developers to sometimes break legacy paradigms when it makes sense. PAF's user interface design is a direct computer re-implementation of the original workflow of 8.5x11" four-generation family groupsheets, and also grew out of the limitations of DOS text-mode displays available at PAF's initial release in 1983. This UI paradigm can certainly stand to be revisited and explored. Furthermore, increasing interaction between users and developers can help bring software bugs to the developer's awareness. Many a developer has been floored to come across the incredibly creative and incredibly obtuse solutions that some users have come up with to working around software bugs or limitations—bugs that could have been quickly fixed by the developers if only the users had an easy way of communicating the existence of the bug to them, and if only the users knew they didn't have to accept broken things the way they are! Developers can't fix bug that they don't know exist, and they're often not motivated to fix known bugs before they know the true cost of a bug in lost productivity.

#2. Centralization

In spite of the overlap with other research communities that are out there, the exact combination of elements present in FHT research is quite unique and especially broad. Nevertheless we are all working towards solving a fairly well-definable set of specific problems. Furthermore, these problems only really need solving once really well, as long as that solution is genuinely useful, flexible, accurate and accessible to all. Additionally, the human resources in any one subproblem area of FHT research have so far always been very

limited. This argues for a strong need for centralization in the coordination of the development of family history technologies.

To avoid the re-invention of wheels, and to catalyze synergy, we need to make previous work really easy to find—there needs to be a **one-stop shop for all things relating to family history technology research.** With the creation of the RootsTech conference, the LDS Church (specifically, FamilySearch.org) is now uniquely positioned to not only facilitate the centralized curation of family history records, but also efforts to develop the technologies that allow those records to be produced and linked. RootsTech should aim to become (and, it seems, is aiming to become) not only the conference that brings together researchers once a year to discuss tech-related genealogical topics, but ***the* launchpad, hub and hive of activity** that everybody in the community uses to connect and collaborate on finding solutions to common problems, not only at the RootsTech conference but throughout the year between conferences.

The next three points describe the infrastructure necessary for rootstech.familysearch.org to become the one-stop shop of FHT: mailing lists, a knowledgebase and source code repositories.

#3. Communication (mailing lists)

The most important step in building any community is establishing the community's means of communication. The rootstech.familysearch.org website is a great start: under the "Community" tab at the top, the website currently includes links to a social media group ("Join us on Facebook"), solicitation of comments ("We want to hear from you"), a Call For Papers, and a link to a directory of affiliated bloggers who will be reporting on RootsTech 2011. However there is also a need for *symmetric (participatory) dialog* that is *enduring and searchable* in nature: the RootsTech community needs a directory of **open mailing lists or forums**, broken into different topics, with **full searchable archives**, clearly linked from the RootsTech main page.

Note that one particularly important feature for a mailing list's archives to be useful is the ability to moderate posts up or down, and/or the ability to "pin" or star important posts, so that posts containing particularly important and/or relevant information can be featured at the top of a given forum, increasing the signal-to-noise ratio of the forum. (Not all message board software supports this feature.)

#4. Curation of knowledge (knowledgebase / wiki)

To avoid re-inventing wheels, RootsTech needs to establish a centralized **problem/solution knowledgebase** or wiki of some form detailing:

- The big problems in family history technology
- What has been tried, by whom
- **What has worked** *and* **what hasn't worked—**and explanations as to why
- A bibliographic database of relevant research (both within the FHT community, as well as relevant papers in mainstream literature).

(A similar knowledgebase on the topic of conducting genealogical research exists in the form of the FamilySearch wiki[5], but another such wiki needs to exist specifically on the topic of the development of family history technologies.)

This knowledgebase will serve as a threshing-floor for sorting out the wheat from the chaff, so that newcomers to the field can quickly get up to speed on previous research without having to start their background research from scratch—and to avoid reinventing wheels. This knowledgebase will also link to the source code repositories of past research and other projects that may be useful in solving a given problem.

The knowledgebase should be hierarchically broken down into grand challenge problem areas. A quickly-brainstormed and by no means comprehensive hierarchical list of example topics is given below. [FHTW proceedings and CFPs for related conferences like ICDAR would also be useful in building the knowledgebase.]

- Document image processing
    - Background removal
    - Locally-adaptive threshold
    - Bleedthrough removal
    - Image straightening and registration
    - Tabular form processing and structure recognition
- Ribbon scanning
    - Automatic adaptive adjustment of sensor lighting and image sensor settings to improve dynamic range
    - Automatic frame slicing algorithms
    - Scanning hardware design
- Indexing of handwritten document images
    - Handwriting recognition for direct information extraction
    - Handwritten shape recognition for high recall, lower precision (actual recognition done by human)
    - Improvements to machine learning that directly support family history work, such as hierarchical multiscale recognition algorithms (i.e. simultaneous solving of global and local shape recognition tasks), simultaneous recognition and segmentation, biologically-motivated handwriting recognition algorithms, etc.
- Genealogical record ontologies
    - Document structure understanding
    - Family history record structure understanding
    - Automated schema extraction and population
    - Automatic inference of family structure given information overlap and record juxtaposition
    - Ontological inference of missing data
- Data normalization
    - Person name normalization and fuzzy matching
        - Heuristic models of patronymic / matronymic and other naming schemes
        - Name change detection based on genealogical context
    - Placename normalization and handling of historical changes

[5] http://wiki.familysearch.org/

        - Handling of multiple different placename hierarchy schemes / missing placename hierarchy levels
        - Historical name changes
        - Geolocation of historical names
    - Automatic and semiautomatic record matching / merging / linking
    - Handling of disagreement between genealogists: alternative views
- Genetic data
    - Inference of historic gene pools
    - Inference of population and/or identity of unknown ancestors
    - Inference of surname based on Y chromosome
    - The “dark matter” of genetic genealogy: going beyond Y chromosome and mitochondrial DNA lineages—making use of autosomal and X chromosome data for inferring genealogical history
    - Inference of population history: migrations and admixture
- Visualization
    - Alternative visualization methods for pedigree charts and other genealogical data
- Augmentation of genealogical records with anecdotal metadata / rich data types
- Data archive and access in a digital age
    - Keeping one step in front of technological obsolescence
    - Alternative backup methodologies: spinning data on magnetic storage / laser- etching on metal plates / …
    - Image compression issues

#5. Code reuse (source code repositories)

RootsTech needs to host source code repositories for open source FHT projects. These source code repositories should employ a distributed revision control system (ideally “git”, created by Linus Torvalds for collaborative development of the Linux kernel). By offering to host FHT-related code projects at a central location, these projects will achieve higher visibility and hopefully will achieve the critical mass necessary to see continuing incremental improvements by a wide variety of interested parties. In the case of BYU FHT research projects, hosting the source code on rootstech.familysearch.org will enable the lifetime of the code to endure beyond the graduation of the student and the shelving of the student’s thesis in the Harold B. Lee Library.

In theory, reusing code between projects should result in less work: features and algorithms should only ever need to be written properly once. Reusing code should also result in fewer bugs: bugs only need to be fixed once in the shared code. In practice however, code is not reused as much as it could be. Writing good reusable code is hard. If you write code that is simple and does a job well, it may be too specific to be easily re-used, and is often simple enough to rewrite the code each time a variant is needed. If it’s powerful, flexible and generic, it would seem that the code could be reused for a wide variety of purposes[6], but it’s often unwieldy and overly complex, and is not used precisely because it does more than is needed. Also developer ego and individual taste inevitably result in “Not Invented Here syndrome”: because somebody

[6]“Everybody needs a thneed!”—the Once-ler [the Lorax, Dr. Seuss]

else wrote it, it can't possibly be good enough or work well enough or doesn't feel native enough to one's own project to justify reuse. Code is therefore more often rewritten from scratch than reused.

What is clear though is that reusing or improving existing code is hard, but reinventing wheels often ends up being harder. It is also clear that here is far too much replicated work, both inside and outside the Church, both in the developer and user communities. How many times has a name normalization algorithm been written? A date parsing algorithm? A GEDCOM parser? How many times has a placename gazetteer been built? How many times have duplicated records been created in large genealogical databases? How many times have the same names been submitted to LDS temples for proxy ordinance work?

A central source code repository will reduce duplication of effort. To be successful and vital, open source projects need to achieve critical mass:

- We need to increase awareness of existing tools and resources. The knowledgebase / wiki described above will facilitate this by linking to code that is available for attacking each problem.
- There needs to be a push within the community towards reusing and building upon existing code for new research projects, rather than starting from scratch. New students at BYU wishing to conduct research on family history's grand challenges should be encouraged to browse existing projects and look for ways of improving them for their own research.
- Companies should be encouraged to build upon these open source technologies where appropriate, and to contribute back to these projects.
- Companies should also be encouraged to contribute commoditizable code modules back to the open source commons for the health of the FHT software ecosystem, and to help push upwards the lowest common denominator level of abstraction upon which all players are subsequently able to innovate.
- Strong developer communities need to be established (or will hopefully self-organize) around key promising projects. New maintainers will need to take ownership of orphaned projects in the case that the author or previous maintainer loses interest or drops off the radar (e.g. when the BYU student who created the code has moved on)—care must be taken in source code licensing and copyright attribution issues to handle future contingencies like this.
- The LDS Church and other communities should work to recruit talented members who have programming skills to help out with these projects by donating some of their spare time. (The LDS Church is doing this already with iPhone and Android open source projects through the LDSTech website[7], and the fruits of these initial efforts are so far of very high quality.)[8]
- BYU needs to push to turn FHT-related theses into shipping products (via the technology transfer office) that actually go into production, both so the student sees the fruit of their work go live, but also so that several years' hard work isn't just sitting in a bound volume on a shelf, read by maybe four people. It is a travesty when two to four years of somebody's life, as well as the creation of a powerful new software tool, is reduced to little more than an academic exercise—especially when there is a real need for that tool out there somewhere.

---

[7] http://tech.lds.org/

[8] Relatedly, the Church has crowdsourced name extraction from microfilm images for many years, and recently announced a beta test of a brilliant and forward-thinking project, "Simple Acts", for the crowdsourcing of various tasks to which members can donate some of their cognitive surplus: http://simpleacts.lds.org/ (**Note:** The non-programmer human resources available through Simple Acts could still be leveraged by the FHT community for numerous purposes related to the development of family history technologies, for example in producing a training dataset for handwriting recognition.)

- For-profit genealogy companies as well as the LDS Church need to keep tabs on the FHT-related research that is being done at BYU and at other universities, and where appropriate license technology from those institutions (or assist in the software being open sourced). Even better, the LDS Church and/or genealogy companies could directly sponsor academic research, providing scholarships to promising students in exchange for research in family history technology. It can be difficult for graduate students to find funding from outside grants for work in these areas.

The individual FHT projects that are hosted in these shared source code repositories will need **integration** work to be useful—typically FHT projects solve a very specific subproblem and needs to be integrated into a bigger data processing pipeline. Fortunately tools and services for creating pipelines of inhomogeneous software parts have improved dramatically in the last few years (for example, hadoop allows for the easy creation of a MapReduce pipeline of individual processing operations and facilitates running it on a large cluster of computing nodes; SWIG makes it easy to automatically generate bindings for routines in any popular programming language so they can be called from any other popular language; Amazon rents large-scale cluster computing resources on their EC2 service at reasonable rates; etc.).

#6. Commoditization of technology

Beginning several years ago, the IT department of the Church of Jesus Christ of Latter-day Saints began a complete overhaul of their IT strategy. The Church brought on board Jay Verkler, Ransom Love and other progressive-thinking leaders in information technology, hired numerous young new programmers, completely re-tooled internally to use Java and new Web technologies, revamped software development practices etc. Through these efforts, the LDS Church has begun to produce some top-quality public-facing web tools, including the new FamilySearch.org website (formally known internally as “The Common Pedigree”). However it’s really not part of the Church’s mission to become a world’s leading authority in the *algorithms* that are most relevant to family history research, even if it does make sense for the Church to build some of its own infrastructure. And as time goes by, it may make more and more sense for the LDS Church to license technology developed externally and/or sponsor external development of technologies rather than developing solutions in-house.

This is analogous to historical efforts to develop space technologies. In the early days of space exploration, you couldn’t just buy an off-the-shelf rocket engine; only governments with multi-billion dollar budgets could design and build rockets that could put an object in orbit or a man on the moon. However in 2004 a private company, Scaled Composites, won the $10M Ansari X-Prize for putting a civilian in space twice within a two-week period (see “Competition” below). The US government subsequently awarded multiple heavy-launch contracts to SpaceX to replace its own aging shuttle technologies, and recently canceled its own extremely costly Constellation program that was intended to replace the shuttle. We have now reached an era where the US Government no longer builds some of its own key space technologies. It doesn’t need to, it’s far cheaper and more efficient to buy private off-the-shelf technology—and the development of that technology is now in the reach of startups that are tiny compared to the scale of the US Government.

Every industry goes through this transition towards commoditization of technology. It already happened in the PC hardware space (generic PCs), it has long been happening in the operating system space (Linux and BSD), it continues to happen in the web technologies space (Apache, GWT, Firefox/WebKit, HTML5), and is currently happening at an alarming rate in the mobile space (Android / iOS). The interesting thing about

the commoditization of technology is that it affords all players a new solid foundation of common-denominator technology to build upon, and innovation can move up to the next level of abstraction. Nobody builds their own web server anymore, it doesn't make sense to reinvent that wheel when there are incredible commoditized open source options; everybody just uses (and innovates on top of) Apache/Cherokee etc. And in the mobile space, the availability and commoditization of remarkably rich app development platforms such as Android has led to a Cambrian explosion of innovation diversity in the design of mobile apps that simply was not possible before powerful and featureful mobile operating systems became a commodity on the majority of cellphones.

Many companies fear that the commoditization of technology, combined with the open sourcing of commoditized technology, is a recipe for destroying business models and even entire industries. It is understandable that this would be a big concern for the FHT industry, where the market is relatively small and customers often can't pay much for products or services. As a result, some genealogy companies try hard to protect their IP and resist forces that would lead to commoditization[9]. History has shown that companies that understand that commoditization is an inevitable process, disruptive though it is—companies that learn to adapt nimbly to the changing tech landscape and begin building one level above the current level of abstraction—are the very companies that survive and become the next generation of market leaders.[10]

There is a lot of discussion about when or even whether a technology should ever be open sourced. Open source business models typically revolve around expertise and support contracts, however several points should be noted: (1) software that is partially or completely open source can still be sold, sometimes with premium features in the commercial version (depending on licensing); (2) opening the source code of a piece of software can actually help a business in unexpected ways by revitalizing the entire software ecosystem that the company operates within; and (3) ideally, open source software also attracts contributions by others in the community and sometimes even brings in contributions from competitors.

Some situations when opening a project's source code may be helpful could include:

- When the software's general availability is crucial to the basic vitality of the ecosystem in which a company operates. (e.g. Google created and released their own browser as open source, in spite of the fact that their competitors Apple and Microsoft had their own browsers; Google Chrome dramatically upped the ante and as a result of this competition, all browsers became significantly more powerful within 1-2 years. Whatever browser people use, Google still wins!)
- When a piece of software has ceased to give a company a competitive edge but would be immensely useful to others.
- When a company feels it can still stay ahead by innovating on top of the open sourced code, but there is a clear need in the community for the code and it makes sense to share it.
- When a piece of software is being retired but could still prove useful to someone. ("Abandonware")

---

[9]While I was working at SMGF, we had a company come to present to us some amazing new patented technology that they were convinced we would want to buy or license. Their patent was for the ability to zoom in and out of a pedigree chart. I had already implemented pedigree zooming for SMGF, and my reaction to their tightly-guarded secret technology was, "I thought to be patentable, an idea had to be non-obvious?"

[10]See *The Innovator's Dilemma*, Clayton Christensen

#7. Competition

Some in the traditionally *pro bono* genealogy community may be puzzled or disappointed by the growing commercialization of genealogy. Many genealogists are retired and don't have a lot of money to spend on genealogy services[11]. However it is important to recognize that commercialization, in particular the aspect of marketplace competition, is one of the most important and efficient drivers of innovation.

Under new leadership, the LDS Church's IT department has wisely recognized **the need for the creation of viable business models to accelerate the rate of innovation in family history technology**, and has begun to provide API access to the familysearch.org database so that innovation can happen externally. Creation of a competitive marketplace is the right direction to move in to speed up the development of needed technologies. However it may also be wise for FamilySearch to **sponsor competitions** to intentionally incubate competitive development of technology. This has the potential to also result in the creation of new companies.

These two aspects of competition as a driver of innovation are discussed below.

(i) Friendly and healthy competition between players dramatically accelerates innovation

In 1990 the Human Genome Project was announced by the NIH. It was projected to cost $3-10B and would take at least 15 years to sequence one genome. In 1998, Craig Venter (who had left the NIH to form his own company) announced that they would sequence the entire genome in 18 months for only a cost of $300M. The competition between public and private projects kicked both groups into high gear, and the first working draft of the human genome was jointly announced by the two projects in the year 2000. Notably, the competition afforded by Venter's Celera Genomics caused the public consortium's efforts to be completed several years early, and reduced the cost of the project by billions of dollars. By 2010, the cost of sequencing a genome using Illumina's Solexa technology was less than $20,000, representing a halving in sequencing cost every 9 months since 2000.[12] At this continued rate, a complete genome should cost $200 in approximately 5 more years.

As noted above, in creating an API for the new FamilySearch.org database that can be used by corporate value-added providers, the LDS Church has acknowledged the importance of creating a vital for-profit competitive marketplace to drive innovation. This is a wise move. It is worth noting therefore that in spite of the previous observations about the need for collaboration, cohesion and cooperation in the community, that there is also a real need for **healthy and friendly competition** between players in the FHT community.

Competition is important not just to drive innovation, but also because there is a real danger in creating a monoculture. Biology teaches us that diversity is extremely important for disease resistance, lest a single disease wipe out an entire homogeneous population. In the open source world, a complaint that is frequently

---

[11]Fortunately, the "invisible hand of the market" (the self-regulating nature of the marketplace) should take this into account, making genealogy generally affordable in the long run based on how much people are able and willing to pay—nevertheless this still presents an issue to commercialization.

[12]"Illumina Drops Personal Genome Sequencing Price to Below $20,000"—Bio-ITWorld, June 30 2010. http://www.bio-itworld.com/news/06/03/10/Illumina-personal-genome-sequencing-price-drop.html

heard is that there are too many competing projects with similar goals that spread the developer base too thin. “If only all the KDE developers came over to GNOME and worked on our codebase, we’d have twice the manpower and GNOME would be awesome!” Actually the opposite is probably true. People are diverse so the things they want to work on are diverse, and diversity is good for the ecosystem.

Of course the need for competition should be balanced with the need to not senselessly duplicate work in parallel or re-invent wheels. We need to put our individual and combined weight behind a small number of projects that move the state of the art forward in specific directions.

(ii) Incentivization of progress through grand challenge prize competitions delivers tomorrow’s innovations today

The state of the art in technology can be dramatically pushed forward through the creation of specific prize competitions that reward **measurable progress** in grand challenge problem areas. A famous example is the $25,000 prize offered by New York hotel owner Raymond Orteig in 1919 for the first allied aviator to fly nonstop from New York to Paris or vice versa. This prize was won by Charles Lindbergh in 1927 in the Spirit of St. Louis. After his historic flight followed the “Lindbergh Boom” as public interest in air travel took off and air industry stocks soared. This is widely recognized as a significant turning-point in the commercialization and commoditization of flight.

There have been numerous other examples throughout history where a grand challenge prize has pushed the envelope of what was possible, but there has been a surge of interest in grand challenge prize competitions in the last decade. Some of the most famous recent grand challenge prize competitions include:

- The $10M Ansari X-Prize for privatized suborbital space flight with a civilian pilot, won in 2004 by Scaled Composites.
- The DARPA Grand Challenge, a $2M prize awarded in 2005 to Stanford for being the first team to pilot an autonomous vehicle through 175 miles of desert terrain.
- The DARPA Urban Challenge of 2007, with the first place prize of $2M won by Carnegie-Mellon, a second place prize of $1M and a runner-up prize of $500k, was a competition to build an autonomous vehicle that could successfully and safely navigate an urban environment.
- The Netflix Prize, a $1M prize awarded to the first team that could improve the accuracy of Netflix’ movie rating prediction algorithms by 10%. This competition sparked a flurry of activity, with thousands of programmers working into the early hours on evenings and weekends working feverishly to try to beat Netflix’ benchmark. This competition resulted in the formation of hundreds of teams scattered across the globe, and elicited a great deal of discussion in forums about the best ways to solve this problem. The challenge proved asymptotically harder the closer a team came to 10% improvement. The winning team, BellKor’s Pragmatic Chaos, was the amalgamation of three of the top teams—it turns out that by combining the results of each of their best algorithms, the resulting accuracy edged past 10%. (Just hours later, another amalgamation of teams achieved the same feat and came in second.)

There are several important lessons to learn from the success of these grand challenge prize competitions:

- In each case, the winning team accelerated the anticipated arrival of a specific technology by many years.

- It has been shown repeatedly in these grand challenge prize competitions that **the total investment into the competition effort by teams exceeded the total prize money by at least a factor of ten—** for example, the Ansari X-Prize awarded $10M but more than $100M was spent by the teams that were competing. That's an amazing return on investment in the future of privatized space flight.
- Framing the challenge as a competition had the effect of significantly adding to the thrill of the hunt: a great many participants of these challenges stated that they weren't in it for the money.
- The X-Prize Foundation has now overseen six different multimillion dollar grand challenge prize competitions, and is preparing several more for release. Peter Diamandis, founder of the X-Prize Foundation, has summarized the success of these grand challenge prize competitions by saying, **"You get what you incentivize."**

One way to incentivize innovation on a much smaller scale is to run a **software bounty** program. This has typically been used by open source communities to accelerate progress in the dusty cobwebbed corners of open source projects. A user will post online that they are willing to pay a certain amount of money to have a specific bug fixed or a specific feature implemented. Other users can chip in and add to the bounty amount. When the bug is fixed or the feature implemented, the best implementation wins the bounty, or the funds are divvied. Websites exist for creating and managing software bounty projects. Also some high-profile companies like Google and Mozilla have created their own software bounty program, awarding hundreds or thousands of dollars to users who discover and report important security bugs to them. (Note that software bounty rules and conditions should be well thought-through before they are announced, so that the requirements to win the bounty are quantifiable and specific.)

In summary, competition is good for innovation, and it would be prudent for the FHT community to find ways to **incentivize measurable progress**.

#8. Cross-pollination

An important step to building any community, and especially to avoid stagnation in the community, is **outreach** with the purpose of achieving **cross-pollination and infusion of new ideas and talent**. RootsTech, as an organization as well as as individual participants in the RootsTech community, needs to actively try to attract people from other related research communities (perhaps students, through sponsored competitions) who may be interested in the fascinating and difficult grand challenges presented by family history research.

The LDS church has archived in the granite vault in Little Cottonwood Canyon what is probably the world's largest archive of historical document images, and, in the databases of FamilySearch.org, the world's largest database of electronically linked and carefully ground-truthed genealogical records—data that would be extremely hard or expensive to come by otherwise. There *must* be researchers around the world that would give anything for access to datasets like this—if they even knew that they existed! There are entire document image analysis and recognition conferences and journals—for example, ICDAR (the International Conference of Document Analysis and Recognition), IWFHR (the International Workshop on the Frontiers of Handwriting Recognition) and IJDAR (the International Journal of Document Analysis and Recognition) —as well as communities of brilliant scientists that already work in related fields, such as CEDAR (the Center of Excellence for Document Analysis and Recognition at SUNY Buffalo)—and there are numerous smaller research groups around the world that publish in the aforementioned conferences and journals, and

whose interests overlap a lot with the goals of RootsTech and with FHT research in general. There are also large communities involved in population genetics research that could help accelerate progress in genetic genealogy. Why are we not reaching out to all these communities and offering them the chance to work with the huge corpora of data that we have access to, and encouraging them to contribute to the RootsTech conference and community, whether by suggesting research collaborations, sending targeted CFPs, inviting promising researchers outside the normal circle of invitees to attend RootsTech or give a talk about their own related work, etc.? Are we trying to get FHT research published outside the usual circles to raise the visibility of this work and these datasets?

We must not be insular in our approach to solving FHT's grand challenges. The infusion of new talent and fresh minds into the community will dramatically increase the quality of FHT research, and will re-invigorate the efforts of long-time members of the FHTW community.

#9. Consistency

The *grand keys to sustainable growth and progress* in any endeavor have always been **consistency** and **intensity**. These two keys are well known in the field of bodybuilding, as a Google search for those keywords will attest[13]. Consistency is particularly important for progress, because "The ratio of something to nothing is infinite" (—Peter Diamandis, Founder, X-Prize Foundation). Any step forward is a nonzero step forward. And *consistency* in incrementally stepping forward is key to eventual success.

If the RootsTech community rallies together to set concrete, achievable goals and then consistently and determinedly works to achieve those goals, progress will be made.

#10. Courage

Repeating once more Joseph Smith's rousing quote about how family history work should be pursued: "Courage...and on, on to the victory!" What is it that specifically requires courage in the development of family history technologies?

- We need courage to move forward to try to solve hard problems, even if we don't quite know how to articulate them or don't know where to start.[14]
- We need to have the courage to try to build upon and improve existing code where possible, to avoid reinventing wheels.
- We need to have the courage as individuals and as companies, where appropriate, to contribute back to the commons for the good of the whole, trusting that prudently doing so will not destroy important business models if innovation keeps happening on top of the shared platform.

---

[13]The principles of *consistency* and *intensity* were also the specific key approaches to scripture study employed by President Joseph Fielding Smith, tenth President of the Church of Jesus Christ of Latter-day Saints, and one of the best-known scriptorians in the history of the Church (as retold by his grandson, Joseph Fielding McConkie).

[14]"Courage is closely aligned with faith—the moral imperative that we would not have been given this work to do or these problems to solve, if there was not a way." —Bill Barrett

- Companies will hopefully find the courage to contribute some funds towards student scholarships for relevant basic research, the result of which may or may not be productizable.
- We need to have the courage to figure out how to work together as a community despite our huge range of backgrounds.
- We need to have the courage and determination to find time to work on family history technology in an increasingly busy and distracting world.

And, for those who believe in a moral imperative to search out our ancestors, we need to have courage to hasten the work[15].

Conclusion

Information technology provides incredible tools for family history research. The RootsTech conference brings together an incredibly diverse and talented group of users and developers of family history technologies. Working together as a community to develop new family history tools presents interesting challenges and opportunities. Greater cohesion and a strong online community infrastructure is needed to speed up the sluggish progress of innovation in family history technology, as well as to reduce duplicated effort. This paper has called for a centralized knowledgebase/wiki, mailing lists and source code repository of relevant family history technologies and an effort from all interested parties to build a strong community around that hub. This paper has also outlined the need for both collaboration and competition in the development of new tools for genealogy, and has suggested the calculated creation of prize competitions to incentivize innovation, motivated out of the immense success of historical and contemporary grand challenge competitions. This paper has pointed out key opportunities and potential resources that RootsTech participants may leverage of to build a vibrant community and to maximally accelerate the progress of family history technology.

---

[15]In his final public talk, President Hunter also stated, *"With regard to temple and family history work, I have one overriding message: **This work must hasten.** The work waiting to be done is staggering and escapes human comprehension. Last year we performed proxy temple endowments for about five and a half million persons, but during that year about fifty million persons died. This might suggest futility in the work that lies before us, but we cannot think of futility. Surely the Lord will support us if we use our best efforts in carrying out the commandment to do family history research and temple work. The great work of the temples and all that supports it must expand. It is imperative!"—Howard W. Hunter, "We Have a Work to Do," Ensign, Mar. 1995, 64 (emphasis added).*