**John Quackenbush**

**1/31/2020**

Chair of Biostatistics at HSPH

Genospace – Sold a few years ago

More recently, predictive methods for machine learning

CD: Scope of the committee is to probe the experience of people who have had success in maintaining open source software

1. WebMeV
2. Genospace
* JQ: WebMeV – more typical academic route
* Started about 20 years ago, in early microarray days.
	+ Desktop tool for hierarchical and k-means clustering. More capabilities over the years. Initial funded through a colon cancer grant. Piggy-backed on other funding.
	+ Then funding through NHGRI. History of funding from here and there. Always challenging. Sometimes gaps in funding, but always a lot of interest from the community. Constant requests from the community. Helped make the case to whoever was funding.
	+ NCI, NLM, NHLBI, and others.
	+ 6-7 yrs ago realized the need to convert to a web-based tool. Initially from an R01, rolled into the ITCR program. After a couple of iterations, funded yet again. Tough to manage from a staffing perspective. But for the next few years, support should be robust.
	+ Working with the journal Cancer Research to encourage publication of cancer focused tools. Outreach is important.
	+ Now trying to organize a workshop, all day affair
* GH: how do you transition out of grant support for long term sustainment?
* JQ: have seen a lot of different things. Genospace founded in 2012. Starting point was a stimulus grant we had received. In 2005, asked by DFCI to “solve the bioinformatics problem”. Came up with the idea of creating a core service for “collaborative consulting”.
* Recognized there were a lot of resources that could be creatively linked together. Got a 2 year Oracle grant to link clinical and genomic data. Essentially an internal project to work with them. Learned that Oracle was not the right solution for storing data. Also learned they needed an interface layer.
* Challenge getting buy-in from Data Farber. Wanted to allow clinical scientist and bench biologist to ask questions of data. Surprisingly, institution asked “how do we limit what questions they ask?”
* Through ARRA funds, started supporting large scale genomic analysis. Clinical data plus multi-omic analysis. In 3 years, moved from Oracle to MongoDB and built a model layer and interface layer. Knew the interface had to be really simple but have multiple entry points. Essentially had 6 ways in. Good system but needed further development. Died a political death.
* Got funding from the Adelson Foundation. System continued to evolve. But still, never got institutional traction.
* Decided to build this ourselves. Bought space on AWS and built the codestack from the ground up. Created a core database built around MongoDB, highly indexed. But not perfect and not enough support for relationships. Put a graph data model on top of this and interface layers that draw on the data models.
* Building a nice-to-have tool. Clinicians at DFCI loved it. But, had to turn it into a need-to-have tool. Found MMRF who needed such a tool. They were starting the CoMMpass project. They had spent 2 years with a consulting company, no results. Put in a head-to-head competition with the vendor. In 3 months delivered a working portal and were awarded a contract. Thus, became a contract service company.
* Wanted to collect RWE data from patients. Built a research gate that returned data back to the patients.
* Worked with a diagnostic testing lab who also wanted reporting. Through this, built several database layers.
* But never got venture funding. Got some funding from Thomson Reuters. Biggest source of funds from clinical reporting.
* In 2012, CMS changed their reporting and there was a lot of consolidation in the diagnostic testing sector. Hurt their income.
* Worked with a group – Sarah Cannon - doing clinical trial matching and rapidly built a clinical trials matching system. Ingested 300 studies and associated medical records. Very sophisticated. Also a “Tinder” app for clinical trial matching. Based on the success of this project, sold the company in 2017. This tool met a mission-critical need for the company. Made a nice profit.
* Code had to be proprietary, no longer open source. They are now marketing it to others.
* CD: Very interesting journey! Grants, foundation money, out of pocket – success based on your own dedication and interest in getting things done. Is this path replicable for others?
* JQ: Yes, I think it is. Could easily commercialize many of the ITCR tools. Tried this with some of our radiomics tools, but no success yet. Wish we could take all of the tools we in ITCR are collectively developing and make them available.
* CD: Really like your dedication. In the end, progress is made by a few very dedicated people.
* JQ: Company offices were Starbucks at 6:00am. Exciting and challenging.
* CD: What group effort might be helpful?
* JQ: SBIR grants?
* GH: Got the Phase I, but didn’t get the Phase II.
* JQ: What do you think barred you from getting funding?
* GH: What happens if you don’t get the license out from Partners? Impression that the software was too similar.
* JQ: Company once got an SBIR grant from NIDA for doing queries across the biobank. Also didn’t end up getting the Phase II due to commercialization plan.
* GH: Have done STTR grants that were more successful.
* AR: What was the point of maturation to have the first commercial discussion? Also, for people who don’t have connections going in, how do you contact the marketplace?
* JQ: Entre into Sarah Cannon came through one of the first customers. Challenge was being a small company, competing with SAP, Oracle, etc. Taking on a small company is risky and Sarah Cannon kept putting them off. But after other options failed a few times, took them on as a risk. Tried to talk to venture funders and bounced between IT and healthcare. Need to learn to talk about what they were doing. Tried very hard to market – hard to find the right person. Had a few false starts. Trouble getting traction when they weren’t fitting a mission critical need. Would want to think about this consideration ahead of time.
* CD: How can you keep your software open source?
* JQ: ITCR for now….this is a struggle we all face. Bioconductor model works because it is so disparate. Middleware is hard to sell. One experiment we are trying is charging people to run analyses. Trying this as a way to be sustainable. Working reasonably well – turnkey operation.