

## Case Study: TAIR and Phoenix Bioinformatics

Prepared for the ESA Workshop on *Creating and Implementing Sustainability Plans for Data Repositories*<sup>1</sup>

Nancy Maron, BlueSky to BluePrint

### Background

TAIR, The Arabidopsis Information Resource, is an independent, not-for-profit organization that hosts and curates a [database](#) of genetic and [molecular biology data](#) for the model higher plant *Arabidopsis thaliana*. It began as a grant-funded project created in 1999 by researchers at the Carnegie Institution for Science with funding from the National Science Foundation. Over the course of two grant cycles, Sue Rhee, Eva Huala and their colleagues developed and built up the database.

When the second grant came to an end in 2009, there was a perfect storm of conditions brewing, though they didn't see it at the time: An explosion in the amount of data being generated in the field, due to the advances of genome sequencing and in projects funded to house the data was putting some strain on funders. At the same time, driven by the need to add new features to secure new rounds of grant funding, TAIR had grown more costly to support. And funders, recognizing the burden of indefinitely supporting successful projects that had reached maturity, were reluctant to support operating costs, much preferring to direct scarce funding to new projects.

### Sustainability history

The real moment of truth came when NSF declared that their next round would step down the funding, over the span of 4 years, cutting back 25% each year. According to Huala, "This was alarming and upsetting to us." They grappled with ideas about how to anticipate the future loss in grant funding. "We were shocked because TAIR was so popular and successful. Our board and the community were upset."

They began with a community letter writing campaign, which "generated heat, but no results (and may have angered NSF)." After which, "we started to cut parts of the program and did not see a route forward. "

The core TAIR team started researching possible directions. "We explored the idea from all angles... we scheduled meetings with people who could help us think about this. We talked with people running startups, VCs, different people who had done kinds of projects we thought were relevant." This exploration gave them good grounding in thinking about the paths that might best fit their style and circumstances.

---

<sup>1</sup> Information in this article is based on interviews and email exchanges with Eva Huala, Executive Director, Phoenix Bioinformatics, as well as documentation provided by Huala and available online.

## Organizational considerations

**Go commercial?** They considered the idea of “going commercial – starting a for-profit company, with TAIR as its product. But stepped back from that, concerned it would lead to “serving the highest paying customers” at the expense of the scholarly community they intended to serve. They took warning from the example of the Yeast Protein Database, which had tried that route. As Huala reported, it ended up an expensive product that mainly served researchers at companies, which then encouraged a new grant-funded initiative to take its place in the academic community.

### **Growing bigger – spin out, or stay home?**

They knew that there would be real obstacles in trying to have Carnegie as a host, should they opt to try a user-fee based model. “There is so much financial work that needs to happen; they were not suited to that, and wouldn’t be able to handle the workload. We needed to separate from them and create something new apart from our host institution.” Eva describes how her team considered the risk involved in making leap from comfortable hosted setting to independence: “The risk didn’t matter, since the alternative was ‘throw in the towel and find other jobs.’”

Creating a separate institution, a not-for-profit organization, involved some immediate steps. They obtained permission from the Carnegie IP lawyer to use TAIR and its IP as the basis for this new organization. The NSF allowed them to roll over some funds as a no-cost extension. They rented temporary offices in Palo Alto, ultimately finding “a building slated for demolition.”

## Understanding audience

As they considered which support models might be the best fit, the TAIR team narrowed down the options to either seeking “voluntary support” in the form of donations from companies, libraries and others; or charging for either access to the data or for the benefit of submitting the data.

Prior to 2013, they had already tried seeking support from companies (\$40K from big and \$10K from smaller ones) She described this phase as “a struggle, months of negotiating, and declining responses.” After a few years of working on this model, it was clear that appealing to the good will of sponsors for voluntary contributions would not be enough. What they did learn, though, was that while voluntary contributions were hard to come by, “companies would be open to paying an annual subscription fee.” Some told TAIR that it would be easier to support them, had the request been framed as a subscription fee.

Next, they explored various fee-based models. They knew a “pay to submit” model would “be tough, since we were already struggling to get them to deposit, when it was *free*. “Without a funder mandate to deposit data, the “case for researchers was not strong enough then.” The other alternative, then, would be “pay to use the data.” They saw this as the “simplest route in many ways. Since we had time pressure and it seemed most straight-forward, we opted for that.”

To determine how this might work, the TAIR team began with a careful examination of the existing user base, to gain a deeper understanding of who potential customers might be. “I spent time understanding who was using the database.” She used Google Analytics, which they’d had since 2007, and data from

“some registered users.” This permitted them to see which universities were regularly accessing TAIR by looking at the names of networks, which were often (not always) based on the name of their university.

They did a market assessment to get a sense of how many potential users there might be. The developed models, including assumptions about what percentage would pay and where the money would come from. At first, they thought it might be from individual researchers, but for a few reasons, the focus shifted to libraries. First, individuals did not have the budget to spend, while libraries “accepted the idea in stride.” As for TAIR, an institutional model represented a real savings for them, since there would be fewer customers to chase.

### **Testing the model**

TAIR leadership took several practical steps, as they worked out their first pricing model. They:

- Looked at comparable commercial products to see what they were charging.
- Had a contact in a company who was willing to review their pricing model.
- Developed a tiered model for institutions, starting with the price individual users were willing to pay and establishing tiers, based on usage and assumptions about what rate would equal a little less than the sum of the projected individual subscriptions.
- Sought feedback from university librarians (their future customers). “We said, ‘here is the price,’ and waited to hear the reaction.” Some visits were in person, others were by phone.
- Sought feedback from individual researchers, including conducting a survey to test prices.

### *Making the transition from free to fee*

Navigating the financial needs of the organization with its mission of serving researchers meant thinking hard about what might be free and what might be fee-based. Huala points out that for starters, “all data and software developed up until the point we switched over is still open and free. Our mission was to support the scientists. We needed to balance the financial needs with needs of the researchers, especially those who could not subscribe.”

They arrived at a plan that continues to offer ways to access the content for free, “albeit less conveniently.” They release data freely after it is in the database for one year.

- A metered access approach. Users get a certain number of free pages (not downloads) each month, after which a warning tells them “now you need to subscribe.” This openness enables some level of use for those researchers who may need one piece of information. “We want to enable this kind of use and we know these people will not be subscribers. And it brings in new customers.”
- How did they know where to draw that line?
  - With funding from Sloan, the TAIR team built the software platform to control the subscription options. They first set the metering at 10,000, which was too high to “catch” any users, but “allowed us to test our software in other ways.” They ramped it down slowly, first to 1,000 pages per month. “This caught a few heavy users and we used them to test the subscription software and iron out bugs.” They gradually ramped down further and further, to increase subscriptions. Today, it is set at 75 page views per month but subject to further changes as needed.

**Any pushback?** “A little, but people knew we were running out of funding,” and so there was lots of understanding that there really was not a choice. “Overwhelmingly, people were thankful we were able to keep TAIR going.”

### *Implementing the model*

In 2014 when the model went into effect, TAIR had to take some chances; not all current users might go along with their plan. They informed companies that they would need to subscribe by January 1, 2014 and academic institutions would need to subscribe by April 1. With only the most basic access control methods in place, they announced this requirement, and were pleased to see a few companies and academic institutions sign on. This bought them some time, in which to continue to build the base of subscribers, and better access control software.

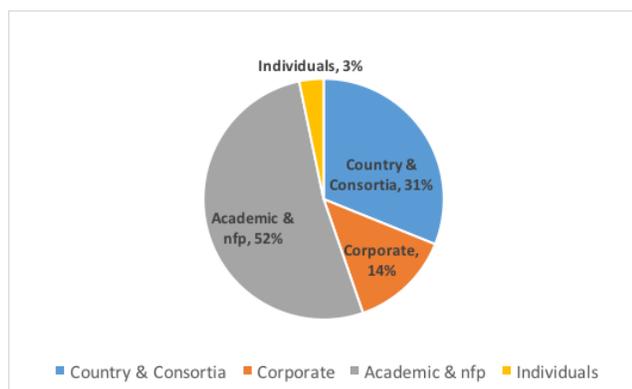
As Huala points out, the flip side of implementing a revenue-generating model is understanding the costs it will take to put in motion. As Huala puts it, “The money is not free.” Among the costs to take into account are the cost of infrastructure, sales, marketing. “We needed to keep those expenses down,” she points out, “to maximize money flowing to (support) the database.”

Their current costs include a sales person, who helps seek new subscribers and handle the renewals and communications with customers, including usage statistics reporting. By analyzing usage data, sales staff can identify potential new customers: when researchers request access because they have reached a limit of free use, they do have an option to immediately buy an individual subscription, but sales staff also are able to “help negotiate an agreement with your university.”

As they have grown, they have continued to refine and modify the model. One example is their pricing model for country-wide licenses, currently in place for both China and Switzerland. A certain amount of work and persistence was needed to find the right contacts to make this happen. In the case of the license with China, one of her team members started by talking with some professors there. Through these contacts, they identified the government agency that could help; after several Skype calls and a coordinated letter writing campaign from supporters, they signed an annual agreement offering access to all Chinese academic institutions.

### Sustainability Model Today

Today, TAIR is supported by over \$1 million annually in contributions from users who pay for access to the database. Thanks to carefully managing costs, they were “profitable in the first year. Starting with a small team and scaling up slowly meant that “there was a period when data and tools were not getting updated at the previous rate, until we had the revenue to support that level of effort.” As of today, in addition to the two country-wide licenses, TAIR has participation from 10 companies, 195 academic institutions and 4 academic consortia.



#### TAIR Subscription Revenue (FY2016)

Country & Consortia	\$321,874
Corporate	\$140,619
Academic & nfp	\$538,541
Individuals & Labs	\$33,817
	<b>\$1,034,851</b>

## **A Q&A with Executive Director Eva Huala, Phoenix Bioinformatics**

Phoenix Bioinformatics is a 501(c)3, registered not-for-profit, and an important part of the sustainability strategy of TAIR. The interview with founder Eva Huala is included below to offer additional context for this organization and its role in supporting TAIR and other data repositories.

*Phoenix Bioinformatics was founded in 2013 by the staff of TAIR, a curated database for plant genome information. After TAIR lost grant funding we pioneered a new sustainable funding model that provides support for TAIR. Our nonprofit mission is to help other projects achieve sustainable support using the tools and expertise we developed for TAIR.*

*(from <http://phoenixbioinformatics.org>)*

### **How do you identify new clients?**

We give talks about our Phoenix Bioinformatics work at conferences such as Biocuration where staff of databases are likely to attend, put up Phoenix Bioinformatics booths at other conferences, occasionally publish our work, and talk about it to whoever is interested. We have also 'cold called' some resources but that has not been very effective for us. I think most people hear of us through word of mouth.

### **What is your customer base? Is it similar for different data repositories? Or is new customer discovery needed each time?**

It's somewhat different but there's lots of overlap with regard to institutions. Little or no overlap on the individual researcher level. Our first partner, BioCyc, is used by more medical research institutions so there were some in that set that had not subscribed to TAIR, which is used by basic and applied plant research programs. So far, each new partner database has been well-established, with its own customer base. We use Google Analytics and any registration data to figure out which academic institutions or companies are heavy users of the resource and we start with those.

### **What other services does Phoenix Bioinformatics offer, aside from distribution?**

1) We wrap our paywall services around the existing data resource without altering it in any way. The way this is done is to send all the web traffic to our paywall first, and then use the rules we've built within it to forward the web traffic on to the data resource or show subscription pages depending on the user's subscription status and number of free views the user has had for the current month. 2) We handle the sales, license negotiation, invoicing and financial tracking. 3) We support the librarians in various ways after the sale, including reporting usage to them and updating the IP range of the institution in our paywall when that changes. 4) We provide automated options for individuals to subscribe and pay via credit card.

### **What is the overall cost/revenue structure for PB? In other words, its own sustainability logic?**

Partner databases share a portion of their subscription revenue with us to support Phoenix operations. The long term financial goal is to use our share of revenue from existing partners to cover the cost of bringing on new partners. In order to get through the initial period when we are bringing on new partners without much of a revenue base to support it, we applied for and received a two year Sloan Foundation grant that supports the onboarding costs of the first few partners. Once we have several partners on board, we believe the revenue from those will support the effort involved in bringing on new partners and the system will be self-sustaining.

**Do you see other opportunities for services to support data repository sustainability, so that each PI need not play every role from scientist to developer to marketing director?**

Yes, definitely! Sometimes we see that the technical choices made by developers of new DRs are not ideal for long term maintenance of a resource, and there are also cases where users are not brought in to provide input on the user interface design, to make sure the DR is intuitive and serves the needs of the intended user community. We hope to someday serve as a scientific/technical consulting service and hosting environment where new DRs can be developed with help from our software engineers and curators. We think we can help new resources develop in a way that will make them stronger technically, scientifically and financially. We can also do some integration across resources to make discoverability easier.

###