Eduardo's ideas for reinventing scientific publishing (2016)

I've been thinking about these things on my own for a while, and finally got to write them down. I would be very keen to feedback and discuss them. In reshaping scientific publishing, below are the most important issues I believe should be addressed, which are artifacts from our current publishing model.

No reproducibility and feedback mechanisms. ([article from Nature](#))

A staggering amount of publications don't provide enough information to make their results reproducible in an unambiguous manner, or even re-derive their conclusions because they don't provide access to sufficient data. This is in large part due to the arbitrary restrictions on the amount of information that publishers will accept, and the fact that authors are often unconcerned about making additional data available elsewhere.

Additionally, there are almost no feedback mechanisms other than retractions, meaning that a lot of publications fall in the gray area of "having a serious issue but probably not enough to warrant a retraction". The community has no tools to address these issues and authors and editors are often unwilling to, not to mention the effort it takes and the very slow pace of such developments.

No way to publish unsexy results, in particular negative results and single observations without a story behind them.

You don't even need to point to people that not being able to publish negative results is an issue, but a more subtle one is that of publishing single observations. If you come across a consistent and unexplained observation during your studies, but cannot afford to investigate it in depth, it will likely be shelved and no one other than lab-mates will hear of it. We have no way of socializing such observations, which can even be widespread and be the tip of a larger uninvestigated phenomena.

No way to discuss work in progress, investigations are kept secret and can take a very long time to share results.

Because of our impact factor, prestige centered mentality and the paranoia that people will steal each other's ideas we often act in the very opposite way of how we should do science. Science is (or should be) a social enterprise! And science should be open source! Yet to an unbelievable degree, unless you personally know the people involved, in many narrow fields it's impossible to know what is going on, because what's happening today will only be shared 5 years from now when some post-doc publishes his results.

You know, the open-source community probably has an even more complex and developed ego-system than the sciences, yet they move at breakneck-speeds, where projects make available nightly builds (the software version including the previous day modifications). Why such stark contrast with the sciences, if both systems are ego-driven? My hypothesis is that the slow pace of the sciences is an artifact of the ingrained mentality from when we had actual constraints to the speeds with which we could disseminate information. The constraints are long gone (and indeed the vast majority of the open source community has never come to know them), but in the sciences, where seniority plays a large role, the habits created by those constraints are still largely in place.

Here's my idea for addressing these issues:

I think the most efficient format for science sharing would be as blog-like entries, which an investigator could publish as soon as he has anything he deems worthy of sharing. This would be quite akin to open source contributions, where developers commit their changes as soon as they have a working copy of what they are working on. Because all changes are tracked and logged, everybody has their rightful claim to credit, and indeed the number of commits in a developer's GitHub profile means a LOT, it's one of the first things an employer will look at. If we could create a "GitHub for science", restructuring things so as to offer a science publishing analogue to the commits in software development, creating a system where scientists have an incentive to share things as soon as they can, I bet efficiency would improve by at least an order of magnitude.

More than that, if we can create a platform that offers solutions to all the "information management" issues in a scientist's life, that would be the killer application for science. So we're talking about a platform that can serve for data storage, lab management, lab notebook, blog-style entries about the pace of your projects, reports on isolated observations, with proper DOI indexing, allowing citations and revisions, and especially providing a good platform for community discussions and reviews. The nearly insurmountable challenge which will likely take generations is to convince researchers to switch from the traditional article publishing model to this kind of continuous self publishing and discussion. Phew.

This is why the current efforts in bringing the self archiving culture into the life sciences are so valuable ([like this](#)). They're paving the way to even biggert things people don't even think about, and they're invaluable because they are eroding the most insurmountable of barriers: culture. Even if such a system as I described existed, user adoption will always be the bottleneck. One way to mitigate adoption concerns and conciliate workings with current practices, is to invite users to outline the progress of their research in the platform, which at some point ends up becoming an article in a traditional magazine, and then the authors can archive the entirety of their data. The more self archiving becomes an accepted practice, the lesser the barrier for getting users into the platform.

Cultural issues aside, I think all the rest of this can be tackled in a startup-like fashion: Build something that users want. At first, try to get a small part of it right for a small set of early adopters, and direct your growth from there: the implementation of new features is guided by feedback of current users and what changes would bring in the most users, with one crucial detail: user retention is critical. In startups, as long as you're growing, you're ok with losing some users by alienating them. Here, because the goal is to embrace the entirety of the scientific community, you can't afford that. It will be better to grow at a slower rate and retain users, than risk getting to a point where a large fraction of users opted out and does not want to come back into the platform.

If everything went unimaginably perfect, the final platform would be something like a cross between GitHub, arXiv, Facebook, ResearchGate, LabGuru, Quartzy. To some extent all these initiatives

capture some aspect of my proposal. Additionally, the platform should also allow for the curation of large databases, like the PDB does. This is fundamental because many areas lack a convenient and standardized repository for data that does not fit into a publication. For example, there is no database for storing and sharing molecular dynamics trajectories. Almost always the authors don't make the simulated trajectories readily available, and indeed it's hard to find molecular dynamics trajectories even for didactic purposes. This is also because those data-sets are usually pretty big, so it's costly to make them conveniently available. Interestingly, this very hurdle provides a sensible avenue for revenue: charge users for the download of very large data-sets. In the case of extremely large sets, the physical media can even be mailed to the users (as Google sometimes does).

Finally, there is one last requirement for such a platform: it has to be completely open source. By that I mean not only that all the information is freely accessible (with the possible exception, as described above, of large data sets), but also that the source code of the platform has to be open. There has to be absolute transparency. If someone thinks they can do things better, they should be able to just grab the code, set up a server and go run things themselves. There can be no margin for any sort of monopolization, or else we risk ending back with the current system.

So yeah, those are my two cents. I believe the first step in such a daunting venture would be mapping the space of who is thinking or doing something in that regard, and see if there's anybody good to join (which would certainly spare a few years of work). Quartzy or Lab Guru are perhaps the two best oriented players in this space (in terms of what they're doing and technical capacity to implement things), but developing this platform would be a huge pivot for them which I have no clue if they'd be interested in. Quartzy seems to have the right personality for it. Lab Guru is owned by Nature Publishing Group, so I would guess they have no interest in hastening the downfall of their own empire. But I could be wrong, and it would be delightful. There's also all the arXiv people, but I don't know the extent of their entrepreneurial drive. Oh, and in the process of writing this I also found this startup, [Kynplex](), which are two Harvard students who just got 100k from the Thiel Fellowship, so that could be interesting too (postscript: Nope. They died.).