

Open Source Research — the Power of Us

Thomas B. Kepler,^A Marc A. Marti-Renom,^B Stephen M. Maurer,^C Arti K. Rai,^D Ginger Taylor,^E and Matthew H. Todd^{F,G}

^A Department of Biostatistics and Bioinformatics, Duke University, Durham NC 27708-0090, USA.

^B Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, University of California, San Francisco CA 94143-2240, USA.

^C Goldman School of Public Policy, University of California, Berkeley CA 94720-7320, USA.

^D School of Law, Duke University, Durham NC 27708-0360, USA.

^E The Synaptic Leap. www.thesynapticleap.org

^F School of Chemistry, University of Sydney, Sydney NSW 2006, Australia.

^G Corresponding author. Email: m.todd@chem.usyd.edu.au

Academic and industrial scientific research operate on powerful and complementary models, consisting of some mix of competitive funding, peer review, and limited inter-laboratory collaboration. Enormous successes have arisen from both models. Yet there are clear failures to deliver results in certain areas, such as the provision of drugs for some of the most prevalent of human diseases. Is there a mechanism of research that is not wholly dependent on funding for its operation nor on traditional peer-reviewed articles for its propagation? Open source methods have delivered tangible benefits in the computer science community. We describe here efforts to extend these principles to science generally, and in particular biomedical research. Open source research holds great promise for solving complex problems in areas where profit-driven research is seen to have failed. We illustrate this with a specific problem in organic chemistry that we think will be solved substantially faster with an open source approach.

Manuscript received: 26 March 2006.

Final version: 29 May 2006.

What is Open Source?

The phrase ‘open source’ originally referred to a community-based approach to software development, most famously the *Linux* operating system. It is now commonly used to describe a variety of collaborative methods in a variety of disciplines. The basis of open source development is that the components of the project (source code in the case of software) are made available to all, may be tinkered with by many independently acting contributors, and recontributed to the larger project. The participants are most often unpaid volunteers who donate their time and expertise for the satisfaction of contributing to the solution of a large, complex problem and the peer-recognition for having done so. The process is strikingly Darwinian. Those contributions which last are successful and influential.

Open source development is not new. The *Iliad* and the *Odyssey* as we know them today were created through countless modifications by generations of anonymous singers and their audiences.^[1] Industry has relied on shared innovation communities since Victorian times.^[2] The modern open source movement started as an effort to defend one

such collaboration (the community that grew up around AT&T’s *Unix*) against corporations wanting to privatize the innovation.^[3] One of the best-known recent successes of open source development has been the *Firefox* web browser.*

The open source paradigm is similar on many levels to the process of conducting research in academia, where solutions to outstanding problems of interest to the larger community are put forward, shared openly through widely accessible journals and ultimately, pending the judgment of many independent minds, incorporated into the shared corpus or forgotten. Conferences, workshops, and ‘invisible colleges’ of researchers facilitate the spread of nebulous concepts, unpublished results, or tentative ideas. Progress is almost always made in incremental steps, with many abandoned ideas along the way. Community-wide databases are another similarity. Like modern open source collaborations, these worldwide projects channel the energy and expertise of hundreds of volunteer contributors, tend to be coordinated by a small central group, and (these days) take place almost entirely online. Examples are the high energy physics (PDG) database,[†] and the BiOS project in biology.[‡] Conversely,

* Descriptions of open source successes may be found on the open source encyclopedia *Wikipedia* at en.wikipedia.org/wiki/Open_source

† pdg.lbl.gov

‡ www.bios.org

chemistry relies primarily on private databases. Significantly, many collaborations use their data to do science by predicting values that have not yet been measured. Such predictions are a short step from using *in silico* methods to develop, say, a novel drug candidate or new synthesis steps for an existing compound.^[4]

In several important ways, however, academic research differs from the open source model:

- (1) Open source development often occurs in the absence of funding for the problem being addressed; scientific research is strongly dependent on external funding. This is not to say that a lack of funding means a project is open source, nor that scientific projects only discover what they are funded to discover.
- (2) Results are made available virtually instantaneously in an open source project, whereas publication in an academic setting can take months to years. Indeed, there
- (3) Progress in open source development is usually communicated through open-access, freely accessible channels; academic research is often disseminated through subscription-only journals (though a trend toward open-access science journals such as *PLoS: Medicine* and the *Beilstein Journal of Organic Chemistry* is gaining momentum).
- (4) A scientific paper will report a result of a certain minimum significance (a reflection of that journal's Impact Factor). Open source development can obviously do the same, but results of minor significance are also shared: Open source research is more smoothly incremental, even if many of these increments do not form part of the eventual solution to the problem.
- (5) Open source research, perhaps unexpectedly, employs a formal structure that involves team leaders who assemble



Thomas Kepler received his Ph.D. in physics from Brandeis University in 1989, and did postdoctoral work in neurobiology and immunology. He develops and applies theory and computational tools to study the evolution of the immune system, spatial organization in the immune system, and the design of vaccine adjuvants.



Marc A. Marti-Renom received his Ph.D. in biophysics from the Autonomous University of Barcelona in 1999. He was awarded a Burroughs Wellcome Fund postdoctoral fellowship, which he spent at the Laboratory of Molecular Biophysics at the Rockefeller University. In 2003 he was appointed as adjunct assistant professor at the University of California, San Francisco, and from 2006 he will be head of the Structural Genomics Unit at the Price Felipe Research Center in Valencia, Spain. His interests focus on comparative protein structure prediction.



Stephen Maurer was educated at Yale University and Harvard Law School. He is adjunct professor in the Goldman School of Public Policy at the University of California at Berkeley. His research interests include open source, intellectual property rights mechanisms, and science policy.



Arti Rai is a Professor at Duke Law School. She has also taught at Yale Law School and the University of Pennsylvania Law School. Her research focusses on intellectual property and new technologies. She currently has a five-year US National Institute of Health grant to study novel open and collaborative approaches to biomedical research.



Ginger Taylor is the founder and Executive Director of The Synaptic Leap, a non-profit organization. Ginger has over 18 years experience in the software industry. Her recent specialty was enterprise portal applications. She was responsible for launching the PeopleSoft Enterprise Portal and during start-up mode managed product development, product management, quality assurance, and product marketing teams. By starting The Synaptic Leap, Ginger is now using her experience with portals to empower online collaboration amongst biomedical scientists.



Matthew Todd received his Ph.D. in chemistry from the University of Cambridge in 1999. He was a postdoctoral fellow at the University of California, Berkeley, Fellow in Chemistry at New Hall College, Cambridge University, and then Lecturer at Queen Mary, University of London, before taking up his present position in Sydney in 2005. His interests are asymmetric catalysis, synthesis, and chemical biology.

disparate contributions into a canonical version. Small laboratory publication-based science does not.

Open Source Communities

As the internet becomes faster, larger, and more usable, open source communities in science are increasingly migrating to the web. The preprint servers in physics and chemistry were the first expressions of this. More recently, open source groups have appeared that are addressing specific problems. We have recently developed an open source community in biomedical research called The Synaptic Leap (TSL).[¶] Our aim is to help coordinate a dispersed community of researchers in biomedical science for any disease where profit-driven research is failing. In a partnership with the Tropical Disease Initiative^{||} we have begun with pilot projects in malaria and schistosomiasis. Tropical disease drug and vaccine development are traditionally underfunded relative to the suffering they inflict because the return on investment is perceived to be insufficient given that the afflicted are found almost entirely in poor countries. Our hope is that open source methods will have a great impact here.^[6]

For example, one of the projects fostered at TSL concerns the synthesis of the active enantiomer of Praziquantel (Fig. 1), the drug used in the treatment of schistosomiasis worldwide.^[7] Schistosomiasis is one of the most serious of the tropical diseases. Praziquantel is currently synthesized and administered as a racemate, but there are several important reasons why the drug should be given enantiopure. Chief among these reasons is that the dose per pill could then be increased, which decreases the ever-present threat of resistance, but administering enantiopure Praziquantel also reduces ‘drug burden’ and reduces the pill size for infants. The challenge for the community is to develop a method for its synthesis that competes with the current \$US 0.07 per 600 mg pill of the racemate, namely \$US 0.23 per enantiopure gram. Since the catalytic enantioselective synthesis of Praziquantel has not yet been published, this is an academic problem, but here it is more a process chemistry problem.

The open source collaboration works in both armchair and wet-laboratory modes. Contributors (academics, industrialists, students, or the public) may contribute intellectually by posting possible routes to Praziquantel, by suggesting suitable reactions for the asymmetric step, or by sharing their experiences on chemical steps suggested. Those with access to a chemistry laboratory can actually attempt reactions of interest and post results, either as part of spare-time activities or as more formal student projects. Rigorous and accurate (rather than anecdotal) reporting here reduces the effort required in self-policing. Contributors with access to relevant materials (such as intermediates or catalysts) may physically share these with others willing to spend time on the problem. Open source communities thus expand the borders of what is already commonplace within chemistry schools.

The enantioselective synthesis is of limited use unless it can be made viable on a large scale at low cost. The open

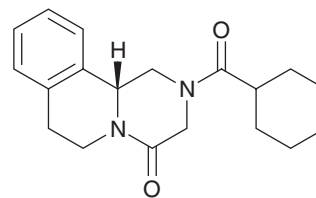


Fig. 1. Praziquantel.

source research model is ideal for this project, since incremental improvements in yields and enantiomeric excesses can make a significant difference to a route’s viability, yet would not *in themselves* justify publication in a journal. The goal is straightforward, and easily quantifiable, so that collaboration among many laboratories around the world can be managed and evaluated appropriately.

The Challenge to Scientific Publishing and Funding

Open source research poses three significant challenges to the traditional operation of science:

- (1) *Peer-Review.* To the extent that open source collaborations discover publishable insights, normal peer-review will continue as before. By their nature, however, open source collaborations will also generate a steady stream of lesser evidence and observations that have not been peer-reviewed. This should not alarm us — indeed, most scientists already ‘surf the web’ for evidence (whether or not peer-reviewed) relevant to their work. How trustworthy are such ‘publications’ likely to be? We are optimistic. First, experience in computing suggests that open source volunteers frequently participate in order to build a reputation — and therefore have a strong interest in self-policing. This dynamic can be enhanced by developing collaboration architectures that emphasize transparency and award honorifics to contributors who demonstrate unusual productivity and insight. Second, volunteer-driven organizations tend to be inherently honest. Blood drawn from volunteers is far less contaminated than blood from paid donors.^[8] Open source projects certainly *want* to succeed but — unlike commercial biotech — they do not *need* to.
- (2) *Publication of Results.* A major question for the journals is: Do we publish papers based on results that have already been published on the web? This is a question that deserves some attention, and we hope that the *Australian Journal of Chemistry* is able to develop guidelines. Many possibilities exist, for example keeping open source results embargoed within the collaboration (in analogy with present-day ‘big physics’ collaborations) for a reasonable length of time. Our opinion is that publication of such results should be perfectly acceptable, because the paper, as it stands, will be peer-reviewed.
- (3) *Grant Funding.* The funding agencies will also need to consider whether they support projects which publish

[¶] www.thesynapticleap.org

^{||} www.tropicaldisease.org

results in an open source format. Agencies clearly encourage the wide dissemination of results, and some are going further: The Myelin Repair Foundation *requires* the grantees to collaborate and coordinate their research.[§] Genuinely open source methods have yet to be addressed within this context. Given the potential power of open source methods, we think that funding agencies will come to see that it is in their interest to promote what is potentially an army of co-investigators to work on a project.

Conclusions

Open source biomedical research is in its infancy. The tools and processes to make it work are evolving. This article is an appeal to the scientific community, funding agencies, and publishers to develop guidelines for how open source research fits into the traditional mode of doing science. We suggest that the funding agencies should channel more resources into open source projects, given the enormous possibilities they generate for worldwide collaborative research. With too much funding open source projects may lose their key strength, participation by volunteers, but the impact of the right level of resources can be substantially magnified through the efforts of those willing to contribute.

Collaborations like TSL are a first step in channeling the enthusiasm of people to become involved, and the power of open source methods to deliver results in important scientific problems will only continue to grow in the coming years.

References

- [1] B. S. Bettelheim, *The Uses of Enchantment: The Meaning and Importance of Fairy Tales* **1976** (Knopf).
- [2] D. Harhoff, J. Henkel, E. von Hippel, *Res. Policy* **2003**, *32*, 1753. doi:10.1016/S0048-7333(03)00061-1
- [3] E. Raymond, *The Cathedral & The Bazaar* **2001** (O'Reilly: Sebastopol, CA).
- [4] S. M. Maurer, *New Institutions for Doing Science: From Databases to Open Source Biology* **2003** (University of Maastricht: Maastricht). www.merit.unimaas.nl/epip/papers/maurer_paper.pdf
- [5] A. K. Rai, *Open and Collaborative Research: A New Model for Biomedicine*, in *Intellectual Property Rights in Frontier Industries: Software and Biotechnology* (Ed. R. W. Hahn) **2005** (AEI-Brookings Press: Washington, DC).
- [6] S. M. Maurer, A. Rai, A. Sali, *PLoS Med.* **2004**, *1*, e56. doi:10.1371/JOURNAL.PMED.0010056
- [7] A. Fenwick, L. Savioli, D. Engels, R. Bergquist, M. H. Todd, *Trend Parasitol.* **2003**, *19*, 509. doi:10.1016/J.PT.2003.09.005
- [8] R. M. Titmuss, *The Gift Relationship: From Human Blood to Social Policy* **1972** (Vintage).

[§] www.myelinrepair.org