

New Institutions for Doing Science: From Databases to Open Source Biology

[Stephen M. Maurer](#)

*Richard & Rhoda Goldman School of Public Policy
University of California at Berkeley*

A paper presented to the
European Policy for Intellectual Property Conference on

Copyright and database protection, patents and research tools, and other challenges to the intellectual property system

Held at the University of Maastricht, The Netherlands, November 24-25, 2003

November 19, 2003

Contact Author: 2632 Hilgard Ave.,
Berkeley, CA. 94709
E-Mail: maurer@econ.berkeley.edu

ABSTRACT

Recently, several authors have suggested that a new method of doing science called “open source biology” is about to emerge. However, very little has been written about how such an institution would differ from existing research institutions. Scientific databases provide a natural model. During the 1990s, scientists experimented with several new database initiatives designed to reconcile private support with the ideals of open science. Despite significant controversy, this paper argues that private/public transactions that unambiguously promote academic science should be encouraged. In principle, research communities can also organize database collaborations to pursue social and political goals. Examples include discouraging software patents, promoting “green” investment, and improving internet security. Finally, the new field of computational genomics blurs the traditional line between database creation and product development. This paper describes how traditional database institutions can be modified and extended to discover pharmaceuticals. The proposed institution (“open source drug discovery”) would be particularly useful for combating Third World diseases. Success would demonstrate that the open source institution is not limited to computer science and can develop products other than software.

1. What Does “Open Source Biology” Mean?

Several authors have recently suggested that a new method of doing science, variously called “open source genomics,” “open source biology,” or “open source biotech” is about to emerge.¹ The idea is intriguing. Although currently confined to computer software, open source methods present an interesting alternative to traditional R&D institutions like intellectual property. So far, however, it is not clear what “open source biology” would actually look like. Articles describing open source biology typically point to (a) computer software written by and for biologists, or (b) projects where biologists publish data but waive intellectual property protection.² The first category surely qualifies as “open source” but is not fundamentally different from collaborations that create non-biology software. The latter category looks uncomfortably like traditional academic publication.³ Somehow, one expects more.⁴

Nevertheless, biology is becoming more and more computer intensive. This fact could make new types of on-line research institutions possible. How should we decide whether such institutions deserve to be called “open source biology?” This paper adopts three criteria. First, the institution should be new, *i.e.*, distinct from existing institutions. Second, in analogy to open source software, the institution should be non-hierarchical, decentralized, and aimed at developing specific products. Finally, the new institution should offer distinct advantages over existing research institutions. These might include (a) fostering collaboration across existing institutional and academic/commercial lines, (b) harnessing the efforts of volunteers up to and including entire research communities, and/or (c) exploiting multiple viewpoints throughout the community.

Earlier commentators have looked to software collaborations like LINUX for clues to what open source biology might look like. This paper argues that academic database collaborations are a better model. Section 2 discusses the “open source” and “open science” concepts and how they relate to one another. Section 3 reviews the tasks that any database collaboration has to accomplish, emphasizing functions that would also be needed for open source biology. Section 4 provides examples of academic database collaborations, emphasizing the strengths and weaknesses of various models. Section 5

¹ D. L. Burk, “Open Source Genomics,” *Boston Univ. Journal of Science and Technology Law* **8**:254 (2002); J. Hope, “Open Source Biotechnology?” (2003), unpublished doctoral thesis available at <http://rssh.anu.edu.au/~janeth/OSBiotech.html>.

² K. Cukier, “Community Property,” *Acumen* **1**:54 (2003); D. Hamilton, “Open to All,” *Wall Street Journal* (May 19, 2003).

³ Biologists seldom patented their discoveries before 1980. S. Maurer, “Promoting and Disseminating Knowledge: The Public/Private Interface,” a report for US National Academy of Sciences (2002), available at http://www7.nationalacademies.org/biso/Maurer_background_paper.html.

⁴ Some scientists have discussed an open source biology model in which researchers – and even teenagers – would trade genetically modified organisms back and forth in the same way that they currently exchange software. In this world, participants would “hack” living organisms instead of code. Hamilton, *supra*. For now, the prospect seems remote.

asks whether recent efforts to build public/private database partnerships are consistent with open science. Section 6 describes future database projects that communities could use to pursue social and political goals beyond simple adherence to open science. Section 7 builds on preceding sections to describe how an open source biology project could be organized to cure Third World diseases. Section 8 presents a brief conclusion.

2. Open Science vs. Open Source

Open Source. In the broadest sense, any software published with human readable source code qualifies as “open source.” In practice, most commentators focus on something much more specific. Legal scholars usually focus on the particular licenses (e.g., GPL, Berkeley Unix) that open source collaborations use to govern how non-members adapt and extend their software.⁵ Social scientists tend to focus on how non-hierarchical, minimally centralized collaborations manage to create a complex product like software. Sub-topics include the incentives that open source collaborations use to motivate volunteers and efforts to test the claim that non-hierarchical organizations produce software more efficiently than traditional corporations.⁶

There is no logical requirement linking licenses such as GPL or Berkeley Unix to non-hierarchical collaborations. Indeed, many traditional corporations already produce and distribute software using open source licenses. Conversely, a non-hierarchical collaboration could easily decide to patent and sell its results. “Open source” and “nonhierarchical organizations” are not logically synonymous, although the same people often advocate both.

Open Science. Open science is variously defined, but tends to connote (a) full, frank, and timely publication of results, (b) absence of intellectual property restrictions, and (c) radically increased pre- and post-publication transparency of data, activities, and deliberations within research groups.⁷ Scholars argue that widespread adoption of these norms would accelerate the rate at which basic science advances.⁸ However, basic science is just one activity. In general, no single institution is likely to be optimal for

⁵ See, e.g., D. Burk, “Open Source Genomics,” *supra*.

⁶ See, e.g., K. Lakhani & R. G. Wolf, “Why Hackers Do What They Do,” *Sloan School of Management Working Paper* 4425-03 (2003), available at <http://ssrn.com/abstract=443040>; J. Lerner & J. Tirole, “Some Simple Economics of Open Source,” *Journal of Industrial Economics* **50**:197 (2002).

⁷ See, e.g., V. Kiernan, “The Open Source Movement Turns Its Eye to Science,” *Journal of Higher Education* (Nov. 5, 1999), available at <http://chronicle.com/free/v46/i11/11a05101.htm>; A. Cottey, “Open Science,” (2000), available at http://www.uea.ac.uk/~c013/open_science/open_science.html.

⁸ P. David, “The Economic Logic of ‘Open Science’ and the Balance Between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer,” (2003), in J. Esanu & P. Uhler, *The Role of Scientific and Technical Data and Information in the Public Domain* (2003), available at <http://www7.nationalacademies.org/biso/>; S. Scotchmer, *Innovation and Incentives*, (MIT Press, forthcoming 2004) at Section 8.6.

every type of R&D that society needs.⁹ Not all science should be open and not all should be patent-driven. The proper boundary between “open” and, for example, “patent-driven” science remains elusive.

For reasons already explained, one can imagine open source biology with or without open science. In some cases, open source biology projects will have little or no contact with commerce, so that open science is unambiguously preferred. However, ambitious projects will often affect the commercial sector as well. Sections 4 and 5 describe instances where relaxing open science rules can potentially accelerate research by tapping private sector resources. Alternatively, Sections 6 and 7 describe cases where a scientific community might take steps to push commercial R&D closer to the open science model. Open source biology collaborations will have to make such decisions case-by-case. In analogy with open source software collaborations, research communities will have to reach near-consensus before proceeding.¹⁰

3. Database Goals.

Many scientific research questions – and even whole disciplines – are impossible without adequate databases. For example, computer simulations of how supernovae make new elements depend on data from thousands of unrelated nuclear physics experiments.¹¹ Similarly, disciplines like genomics and computational biology depend on complete and accurate gene sequences.¹² From a policy standpoint, databases provide a kind of conveyor belt for transferring information from those who acquire it to those who use it. The fact that an experiment has been done – and even published – has little value if it takes inordinate time and effort to locate it. Good databases help society get full value from its investment in science by helping would-be users find, acquire, and understand earlier results.¹³

⁹ S. Maurer & S. Scotchmer, “Procuring Knowledge,” forthcoming 2004 in Gary Libecap, ed., *Advances in the Study of Entrepreneurship, Innovation and Growth*, JAI Press, the Netherlands, available at <http://papers.nber.org/papers/w9903.pdf>.

¹⁰ T. Berners-Lee, *Weaving the Web* (Harper: 2000).

¹¹ Richard Firestone (Lawrence Berkeley National Laboratory), personal communication.

¹² Ben Gutman, “Science Finds its PLOS in the Sun,” *Berkeley Science Review* 5:13 (2003), available at <http://ist-socrates.berkeley.edu:7066/pdf/3.2/policy.pdf>.

¹³ Of course, some proprietary databases also derive value because they contain trade secrets. This aspect of the problem seldom applies to the academic databases under discussion in this paper.

Goals. Good scientific databases share certain features.¹⁴ These include:

- *Coverage.* The value of a database increases with size. Organizing data found in university libraries is only the beginning. Tracking down obscure and/or unpublished results forces the database to tap information scattered throughout the particular research community that produced the knowledge.
- *Curation and Updating.* Good scientific databases recommend best values, predict results that have not yet been measured, and identify suspected errors. These tasks require massive amounts of human judgment. They may also require access to individuals who possess specialized knowledge about particular topics and/or experiments. Both activities benefit from widespread involvement from the relevant research community.
- *Responsiveness.* Databases have no value unless people use them. Databases that fail to deliver the information, formats, and tools that the research community wants are irrelevant.
- *Audience.* Historically, most academic communities made databases for themselves, limiting industry's ability to use and exploit new discoveries. Life sciences companies are particularly interested in obtaining better access to academic data.¹⁵
- *Open Science.* Commercial scientific databases are often expensive and restrict users' ability to copy, redistribute, extract, or modify data. These factors limit the number of users and applications, making commercial products less valuable to society.

In practice, no real database performs all of these functions equally well. Furthermore, grant-supported and commercial databases tend to have different strengths and weaknesses. The next section examines some traditional examples and looks at recent attempts to design hybrid models.

¹⁴ S. Maurer, R. Firestone & C. Scriver, "Science's Neglected Legacy" *Nature* **405**:117 (2000).

¹⁵ S. Maurer, "Coping With Change: Intellectual Property Rights, New Legislation, and the Human Mutations Database Initiative," *Human Mutations* **15**:23 (2000), available at http://www.wiley.com/legacy/products/subject/life/genetics/genetics_humu_mdi.html.

4. Traditional Models and Recent Experiments

Traditional Models. Grant-supported and commercial databases both have a long history. Examples include:

- *Evaluated Nuclear Structure Data File (ENSDF).*¹⁶ This US government-supported nuclear physics database receives and compiles data from a worldwide network of volunteer editors. The resulting database is extremely complete and draws on numerous experts to evaluate results and suggest best values. However, it is hard to use and tends to ignore industrial users.
- *GenBank.*¹⁷ This US government-supported biology database receives massive amounts of gene sequencing data. Community norms and academic journal policies enforce regular deposits by academic and some commercial scientists. However, curation/updating tends to be minimal.¹⁸ Genbank also has minimal software tools. Pharmaceutical companies typically find proprietary databases more useful.
- *GDB.*¹⁹ During the early 1990s, this government-funded database was a leading provider of genome information. However, its leaders were widely criticized for designing formats and tools that ignored community wishes. By mid-decade, it was seldom used.
- *Commercial Bioinformatics Products.*²⁰ Private sector firms frequently sell proprietary versions of Genbank that feature additional curation/updating, user-friendly software tools, and (in most cases) additional proprietary data. These products are noticeably more responsive to users, particularly in the industrial sector, than traditional grant-supported models. Despite extensive price discrimination, many academic biologists cannot afford them. They therefore do a much poorer job of supporting open science.

The foregoing examples show that government- and commercially-supported databases have different strengths and weaknesses. Generally speaking, government-funded databases seem to do a better job of eliciting community-wide information and expertise. They are also almost always available at nominal cost, which favors open science. On the other hand, market forces encourage editors to pay attention to user

¹⁶ <http://www.nndc.bnl.gov/nndc/ensdf/>; See also, National Research Council, *Bits of Power: Issues in Global Access to Scientific Data* (n.a.:1997) at pp. 205-6.

¹⁷ <http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html>

¹⁸ See, e.g., F. Oulette, "Users Must Help to Keep Public Databases Correct," *Nature* **409**:452 (2001).

¹⁹ <http://www.gdb.org/>

²⁰ See generally, S. Maurer, "Promoting and Disseminating Knowledge," *supra*.

needs and preferences. Moreover, commercial databases tend to be better funded. Given that academic and commercial databases are complementary, it is natural to ask whether hybrid institutions are feasible.

Recent Experiments. Most academic databases are understaffed and underfunded.²¹ Over the past five years, biologists have experimented with various transactions that try to reconcile revenue creation with the ideal of open science. Recent examples include:

- *Swiss Prot.*²² In 1996, the Swiss government stopped funding this long-established biology database. Swiss-Prot's founder formed a new company to distribute the database under license. The license provides, *inter alia*, that (a) academic scientists can use Swiss Prot without charge, (b) commercial companies must pay a substantial fee, (c) copies of the database must bear a notice requiring any for-profit user to pay fees, and (d) no user can modify the data without Swiss Prot's permission.
- *GDB.* This formerly-government supported database negotiated a contract with a private partner (SNX) in order to obtain operating funds. Under the proposed agreement, SNX would have received the right to build and sell an upgraded "Private Version" of GDB. In return, GDB would have received funds to support and strengthen its existing, publicly available database. These resources would have included (a) royalties on all net revenues above \$25 million, (b) the right to use SNX's proprietary software for internal, non-commercial purposes, and (c) the right to sublicense SNX's software to academic and non-commercial users of GDB's public domain database.²³
- *The Human Gene Mutation Database (HGMD).*²⁴ This University of Wales collaboration maintains the world's most complete database of published mutations. In July, 2000, HGMD agreed to make a commercial company (Celera) the exclusive distributor of its data. Celera embargoes the data for twelve months before releasing it into the public domain over HGMD's web site. In return, Celera provides unspecified financial support to HGMD.
- *SNP Consortium.*²⁵ In 1998, Perkin-Elmer Corporation entered into a partnership with biologist/entrepreneur Craig Venter to discover and patent

²¹ S. Maurer *et al.*, "Science's Neglected Legacy" *supra*.

²² <http://us.expasy.org/sprot/>

²³ S. Maurer, "Promoting and Disseminating Knowledge, *supra*."

²⁴ <http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html>

²⁵ <http://snp.cshl.org/>; *see also*, M. Morgan, "New Paradigms in Industry: The Single Nucleotide Polymorphism Consortium," in J. Esanu & P. Uhler, *The Role of Scientific and Technical Data and Information in the Public Domain, supra*.

commercially-promising genes. Commentators worried that Venter was about to become “the Bill Gates of the human genome.” In 1999, eleven companies²⁶ joined The Wellcome Trust to form the SNP Consortium, which pays university and academic scientists to discover sequencing data and place it in the public domain. In this instance, industry and science wanted exactly the same thing: Prompt publication without intellectual property protection.

- *Mutations Database Initiative (“MDI”).*²⁷ MDI is a 600-member society dedicated to building better databases for the world’s mutations science community. In September 2000, the author helped MDI negotiate a Memorandum of Understanding in which Incyte Pharmaceutical Company agreed to contribute \$2.3 million over three years so that the community could construct a worldwide database. Under the proposed deal, Incyte would have received the exclusive right to host the database on a commercial web site. This would have let the company use the database to attract traffic to its web site. Academic and commercial users would have been able to download, modify, and extract data for all other purposes.

The foregoing examples show that academic/commercial transactions can generate useful revenue. They also demonstrate that clever transactions can minimize – but not eliminate – restrictions on the use and re-distribution of data. It is not *a priori* clear whether these compromises are consistent with sound public policy. The next section discusses community reactions to recent transactions and suggests criteria for judging future proposals.

5. Should The Research Community Support It?

Except for the SNP Consortium – the rare case in which “open science” coincides with private sector incentives – each of the foregoing experiments has been controversial:

- *Swiss Prot.* Many academic biologists privately argue that the SWISS-PROT model put inappropriate restrictions on users’ ability to extract, modify and re-distribute data. The US government’s National Center for Biotechnology Information (“NCBI”) stopped incorporating SWISS-PROT data in two of its databases because it believed that GenBio’s license would deter private sector companies from using them.²⁸

²⁶ The SNP Consortium’s private sector members include AP Biotech, Astra Zeneca, Aventis, Bayer, Bristol-Myers Squibb, F. Hoffman-LaRoche, Glaxo Wellcome, IBM, Motorola, Novartis, Pfizer, Searl, and SmithKline Beecham.

²⁷ S. Maurer, “Inside the Anticommons: Academic Scientists’ Struggle to Commercialize Human Mutations Data, 1999-2001,” preliminary version available at <http://emlab.berkeley.edu/users/bhhall/ipconf/maurer01.pdf>.

²⁸ J. Ostell (NCBI), personal communication.

- *HGMD*. Many biologists believe that Swiss Prot’s one year embargo is excessive. It is not clear if the community would support a shorter (*e.g.* six month) embargo.²⁹
- *GDB*. GDB’s parent institution fired the database’s director for conducting what it claimed were unauthorized negotiations. It never said whether the proposed deal was acceptable. However, a spokesman has insisted that the institution “might have agreed” if the director “had chosen to discuss his desires with us.”³⁰
- *MDI*. MDI members held an acrimonious debate in October 2000. An NIH representative was present, but declined comment on whether the proposed transaction with Incyte was appropriate. Instead, she suggested that her agency would consider funding the project. Although community members promptly filed a grant proposal, NIH has yet to fund the project as of November 2003.³¹

Community values clearly played an important role in these disputes. However, criticism has almost always been *ad hoc*. For this reason, it provides very few principles for deciding when future agreements might be “acceptable.” One obvious design principle would be to say that an “acceptable” agreement must accelerate science. At least three observations follow:

- *No Other Alternatives*. Any restriction on data is preferable to no data at all. The proposed agreement between MDI and Incyte would have lasted three years. Had it gone forward, the database would now exist and Incyte’s rights under the agreement would be about to expire. At least in retrospect, the mutations community is unambiguously worse off today than it would have been had it accepted Incyte’s proposal.
- *Tangential Applications*. Internet companies frequently deliver information (“content”) free of charge. These models can be readily adopted to provide support for scientific databases. Possible private/public transactions include posting databases on a particular web site as “traffic builders,” selling advertising, selling alert services, and charging for print and CD-ROM versions of on-line databases. Arguably, such restrictions do not conflict with open science so long as the data themselves can be freely downloaded, modified, and reused to do academic and commercial science.

²⁹ R. Cotton (Genomic Disorders Research Centre), personal communication.

³⁰ J. Saunders, “Firings Rock Gene Project at Sick Kids,” *Toronto Globe & Mail* (Nov. 12, 2001); L. Bonetta, “Sackings Leave Gene Database Floundering,” *Nature* **414**:384 (2001).

³¹ S. Maurer, “Inside the Anticommons,” *supra*.

- *Minimal Restrictions.* The most difficult case involves cases where funds are exchanged for supposedly modest restrictions on scientists' ability to use and re-distribute data. Common examples include embargoes, basic *vs.* premium versions of the database, and selling access to academic users at a discount. There is probably no general rule for deciding when such transactions yield a net benefit to science. Instead, communities must decide on a case-by-case basis.

Social scientists – and eventually government funding bodies – can make a major contribution by refining the open science concept so that it provides clear guidance about which transactions are and are not acceptable. Current uncertainty almost certainly has a chilling effect. Greater clarity would let appropriate transactions go forward, accelerating science and saving tax dollars.

6. Database Activism.

Scientific communities share many values. The debates of the late 1990s concentrated on two of them: (a) “advancing the discipline,” and (b) “preserving the principle of open science.” However, these are not the only possibilities. Instead, communities could support initiatives to pursue policy goals in the broader society. Potential “activist” causes include:

- *Software Patents.* Computer scientists frequently complain that software patents are excessive and/or unnecessary. Many blame this condition on the fact that patent examiners cannot find “prior art” demonstrating that particular “inventions” already exist. Computer scientists could fix this problem by building a user-friendly database that summarized prior art.
- *Indigenous Knowledge Patents.* Ecologists frequently complain that multinational companies should not be allowed to patent Third World plants and traditional folk knowledge. As with software patents, better prior art databases would discourage the practice. The American Association for the Advancement of Science has recently funded one such initiative, but more could be done.³²
- *Green Investing.* Many social scientists think that corporations pay too little attention to their impact on society and the environment. Many “green investors” agree, but have no good way to evaluate “responsible” companies. However, many corporate impacts can be quantified.³³ A comprehensive

³² See “Traditional Ecological Knowledge – Prior Art Database” at <http://ip.aaas.org/tekindex.nsf>

³³ Some consulting firms offer independent, third party audits of company behavior. See, e.g., SVT Consulting Home Page, available at www.svtconsulting.com/sra.

database of this information would empower “green investors” and encourage companies to modify their behavior.

- *Cyber-Security.* Computer scientists frequently complain that Microsoft products have poor security. Many blame this condition on a market failure, *viz.* that consumers find it nearly impossible to judge security for themselves. Computer scientists could fix this problem by building a database of known and suspected defects and ranking commercial products accordingly. Posting this data on the Web would provide a powerful incentive for Microsoft and other vendors to improve security.

All of the foregoing projects would create databases. Except for their politics, they are unlikely to look markedly different from the collaborations discussed in Section 4. For this reason, it would be misleading to call such institutions open source science. The next section goes beyond database production to ask whether community collaborations can develop specific products. The ability to deliver products is the hallmark of LINUX-style collaborations and fully justifies the open source label.

7. Using Open Source Biology to Discover Cures for Third World Diseases.³⁴

Background. Until recently, drug discovery was a “wet” science – if scientists wanted to identify potentially therapeutic chemicals, they had to do experiments in test tubes, live cultures, and animals. The completion of various human, disease, and vector (e.g. mosquito) genomes has changed all that. Biologists have learned how to recognize specific genome locations that control metabolism – and identify “targets” that can be manipulated to confer immunity or kill hostile organisms. Researchers also know how to compare targets against large chemical databases to identify potential drugs. Instead of going to the lab, biologists are beginning to do drug discovery *in silico* – that is, sitting in front of their computers.

The rise of *in silico* biology has dramatically lowered the cost of conducting useful drug discovery. Instead of building databases that passively compile knowledge, academic communities can potentially search for medicines. As always, a community-wide collaboration would have to engage widely-held values. Competing with pharmaceutical companies to find cures for Western diseases is unlikely to elicit much enthusiasm. On the other hand, Third World diseases kill millions of people each year. According to the UN, world R&D budgets are an order of magnitude short of what is needed.³⁵

³⁴ This section reports ongoing joint research with Arti Rai (Duke University) and Andrej Sali (University of California at San Francisco). The author is solely responsible for any errors.

³⁵ D. Butler, “Gates Steps Up War On Malaria,” *Nature*, **425**:331 (2003).

Open Source Drug Discovery. The rise of *in silico* biology blurs traditional distinctions between drug discovery, academic database production, and open source. Commonalities include:

- *Making Scientific Judgments.* The ability to “read” databases for targets and drug candidates requires judgment, expertise, and the ability to infer results by comparing multiple data sets. Database editors perform similar tasks when they predict the results of future experiments and infer best values from conflicting experiments.
- *Tapping Community-Wide Knowledge and Expertise.* The ability to recognize a potential drug target depends on ability, knowledge, and experience. As in scientific database production, this fact makes community-wide collaborations desirable. In some cases, unpublished data may show that a particular drug candidate is a dead end³⁶ A community-wide program is the only way to elicit such knowledge.
- *Pursuing a Goal.* Drug development extends the database production model by requiring volunteers to interact in pursuit of a common goal. Open source shows that loose, non-hierarchical groups can work together to perform complex tasks and create specific products.
- *On-Line Collaboration.* Open source collaborations provide a convincing model for how researchers can interact and work together on-line. An open source drug discovery program would almost certainly be built around a Web site where volunteers could examine shared databases. Volunteers would annotate the genome each time they made a discovery and discuss discoveries in chat rooms. The most dedicated and proficient volunteers would eventually become leaders, posting suggestions on which research avenues looked promising or needed workers.
- *Manpower.* Open source collaborations prove that it is possible to extract substantial amounts of labor from unpaid volunteers. The best-known example is probably LINUX. In practice, programmers volunteer for many different reasons including idealism, learning new skills, gaining reputations, and impressing potential employers.³⁷ These same incentives are likely to apply to professional biologists.

Holding Down Drug Costs. By itself, *in silico* biology is not enough. Even if an open source collaboration identifies promising candidates, First World governments and charities will still have to pay for the test tube, animal, and human experiments needed to

³⁶ R. Altman, Stanford University (personal communication).

³⁷ See, e.g., K. Lakhani & R. G. Wolf, “Why Hackers Do What They Do,” *supra*; J. Lerner & J. Tirole, “Some Simple Economics of Open Source,” *supra*.

confirm that the proposed drug is safe and effective. However, this is true of *all* systems of drug discovery. The main advantage of open source is that it is likely to reduce the total life cycle cost needed to get the job done:

- *Discovery.* Open source promises to deliver substantial quantities of unpaid, highly trained manpower. This factor is significant given that private and public Third World drug discovery programs are badly understaffed.
- *Development.* Most current reform proposals try to accelerate drug development by promising to subsidize Third World purchases if and when new drugs are developed.³⁸ In principle, thrifty sponsors should offer a subsidy that barely covers expected R&D costs. In practice, per-drug R&D costs are very poorly known, with published estimates ranging from \$100 to \$500 million.³⁹ Provided that sponsors elicit any discovery at all, they are almost certain to overpay. Open source escapes this trap by placing discoveries in the public domain⁴⁰ so that any company can pursue them. Now sponsors do not have to estimate an appropriate subsidy. Instead, they can let companies bid for the right to conduct drug development.⁴¹
- *Manufacturing.* Because open source drugs would be public domain, any company could manufacture them. Generic drug competition would keep market prices at or near the marginal cost of production.

The failure of Western governments and pharmaceutical companies to cure Third World diseases is almost entirely about cost.⁴² It is reasonable to think that open source biology can break the impasse – and save millions of lives in the process.

³⁸ See, e.g., M. Ganslandt, K. Maskus & E. Wong, “Developing and Distributing Essential Medicines to Poor Countries: The DEFEND Proposal,” *The World Economy* 24:785 (2001); M. Kremer, “A Purchasing Commitment for New Vaccines Part II: Design Issues” in A. Jaffe, J. Lerner, and S. Stern (eds.), *Innovation Policy and the Economy* (MIT 2001); J. Sachs, “Helping the World’s Poorest,” *The Economist* (Aug. 14, 1999).

³⁹ See, e.g., A. Relman & M. Angell, “America’s Other Drug Problem,” *The New Republic* (Dec. 16 2002), pp. 27-41.

⁴⁰ In principle, an open source biology collaboration could stop short of placing drug discoveries in the public domain by attaching “GPL” or other license conditions. This scenario does not materially change the argument.

⁴¹ Generic drug manufacturers and Third World pharmaceutical houses are particularly likely to participate in such a scheme.

⁴² S. Andreopoulos, “Developing Drugs for Parasitic Diseases,” *Science* 300:430 (2003); M. Enserink, “Malaria Researchers Wait for Industry to Join Fight,” *Science* 287:1956 (2000); M. Reich, “The Global Drug Gap,” *Science* 287:1979 (2000); Institute for OneWorld Health, “Fulfilling the Promise of Medicine for the Developing World,” (2003) available at <http://www.oneworldhealth.org/index.html>.

8. Conclusion

Existing academic and commercial models stress competition between individual research teams over community-wide collaborations. It is reasonable to ask whether open source biology can open the door to new research problems that are hard to do under the current system. Existing academic databases show that voluntary, community-wide initiatives are possible. In the information economy, knowledge workers can – and should – be activists who pursue goals beyond open science. The possibilities range from discouraging software patents to finding cures for Third World diseases.