

OPEN SOURCE DRUG DISCOVERY: FINDING A NICHE (OR MAYBE SEVERAL)

Stephen M. Maurer*

I. INTRODUCTION

In retrospect, it was probably inevitable that admirers of open source software production would look for a second act. Before the 1990s, most commentators had imagined that the landscape of innovation consisted almost entirely of patents and copyright. This was never really true – careful policy makers have always considered other options such as prizes, grants, and contract research¹ – but in the closing, patent-obsessed decades of the Twentieth Century the illusion was hard to shake. To observers who had never imagined such a thing, the realization that workers could create new products without central control or private ownership was unexpected.² The fact that open source had occurred in one of society's hardest innovation problems – complex computer operating systems – seemed doubly miraculous. Like moon rockets and integrated circuits, operating systems required far more labor than any single human lifetime could supply. In the Twentieth Century, that meant big, hierarchical teams. It did not really matter whether the engineers were in Santa Monica or Minsk, the long rows of drafting boards and (later) computer terminals looked pretty much the same. The system worked, of course, but anyone who has ever read the comic strip *Dilbert*³ knows that it was not pretty. Suddenly, the open source movement had shown that at least one complex invention – computer software – could be organized in a completely different way. Why not others?

The question remained: What would this second act be? At least intuitively, drug discovery was the likeliest candidate. After all, pills – like software – consist largely of information. Indeed, per-pill research and

* Director, Goldman School IT & Homeland Security Project, University of California at Berkeley and Visiting Professor, Gould School of Law, University of Southern California. I thank Dan Klerman, Ed McCaffery, Arti Rai, Elyn Saks, Andrej Sali, Suzanne Scotchmer, Matthew Todd, Jennifer Urban, Simon Wilke, and the USC Brown Bag Seminar for their many useful comments. I am, of course, responsible for any and all mistakes.

¹ See, e.g., SUZANNE SCOTCHMER, *INNOVATION AND INCENTIVES* (MIT Press 2004) (formal economic analysis comparing the relative merits of intellectual property, prizes, contract research, and other mechanisms).

² In retrospect, open source had many precedents. Shared innovation was common in many Twentieth Century industries. ERIC VON HIPPEL, *DEMOCRATIZING INNOVATION* (MIT Press 2005). Indeed, the process is even older. The literary power of folklore (including, for example, the oral traditions behind *The Iliad* and *The Odyssey*) can be traced to hundreds of retellings in which audiences and entertainers collectively polished an original shared story. BRUNO S. BETTLEHEIM, *THE USES OF ENCHANTMENT: THE MEANING AND IMPORTANCE OF FAIRY TALES* (Alfred A. Knopf, Inc. 1976).

³ See generally SCOTT ADAMS, *WHAT WOULD WALLY DO? A DILBERT TREASURY* (Andrews McMeal 2006).

development expenses are roughly the same as manufacturing costs.⁴ Furthermore, drug discovery and computing seemed to be converging. By the late 1990s, pharmaceutical experts were predicting a new era of rational drug design in which the next generation of pharmaceuticals would be discovered by scientists staring at computer monitors. On the other hand, open source advocates famously argued that having large numbers of volunteers – “many eyeballs” – look at software was a great way to find bugs.⁵ Why not use the same methods to find subtle flaws in the genome? Indeed, the scenario even had a villain: “Big Pharma” with its massively hierarchical teams looked perfect for the part of Microsoft. Finally, biologists were already launching initiatives (most notably the SNP Consortium and the Bermuda Protocol) that renounced patent rights in the human genome.⁶ By 2000, many observers thought that open source drug discovery collaborations were just around the corner.⁷

Nearly a decade later, the revolution has yet to arrive. Some commentators still see a future in which hackers trade open source molecules⁸ and even entire life forms.⁹ Increasingly, however, it is the *absence* of open source drug discovery that needs explaining. Why has so little happened? Most scholars

⁴ Phil Norris, *Virtual Pharma: Re-examining the Impact on Business Performance*, Pharm. Tech. Europe (2003) (estimating that per-drug manufacturing and supply costs total twenty-two percent while research and development costs total twenty percent.).

⁵ See generally ERIC RAYMOND, *THE CATHEDRAL & THE BAZAAR: MUSINGS ON LINUX AND OPEN SOURCE BY AN ACCIDENTAL REVOLUTIONARY* (O'Reilly 1999).

⁶ For a short description of the SNP Consortium and the Bermuda Protocol, see Stephen M. Maurer, *Promoting and Disseminating Knowledge: The Public/Private Interface* (Sept. 6, 2002) (commissioned report for U.S. Nat'l Acad. of Sciences), available at http://www7.nationalacademies.org/biso/Maurer_background_paper.html.

⁷ The term “open source biology” was coined by Molecular Sciences Institute head Roger Brent in 1999. David Cohn, *Open Source Biology Evolves*, WIRED (Jan. 17, 2005), available at <http://www.wired.com/medtech/health/news/2005/01/66289> (last visited July 15, 2007). Human Genome Project scientists and at least one plant biologist seem to have considered using copyleft licenses for data. Janet Hope, *Open Source Biotechnology* (Dec. 2004) (unpublished Ph.D. Dissertation, Australian National University), available at <http://opensource.mit.edu/papers/hope.pdf>; Kenneth Neil Cukier, *Open Source Biotech: Can a Non-Proprietary Approach to Intellectual Property Work in the Life Sciences?* ACUMEN J. LIFE SCI. (2003), extended version available at <http://www.cukier.com/writings/opensourcebiotech.htm> (last visited July 15, 2007). The earliest mention of open source drug discovery in the legal literature seems to be Dan Burk, *Open Source Genomics*, 8 B.U.J. SCI. & TECH. L. 254 (2002) (transcript of April 27, 2001 Symposium on Bioinformatics and Intellectual Property Law).

⁸ Steven Weber, *From Linux to Lipitor: Pharma and the Coming Reconfiguration of Intellectual Property*, in *HOW REVOLUTIONARY WAS THE DIGITAL REVOLUTION: NATIONAL RESPONSES, MARKET TRANSITIONS, AND GLOBAL TECHNOLOGY 217* (John Zysman & Abraham Newman eds., 2006) (“Someone will set up large-scale experiments with open source development processes for therapeutic molecules.”).

⁹ See, e.g., Freeman Dyson, *The Future of Evolution* (2005), available at http://www.metanexus.net/metanexus_online/show_article2.asp?ID=9361 (last visited July 15, 2007) (predicting that housewives and schoolchildren will eventually design and trade life forms), and Rob Carlson, *Intentional Biology*, <http://www.intentionalbiology.org/osb.html> (last visited July 15, 2007) (predicting rise of garage-based “biohacker community”).

offer two explanations. First, open source computing requires very few resources beyond the volunteers' own time. The average pharmaceutical product, on the other hand, costs about \$802 million¹⁰ and includes substantial out-of-pocket costs for chemical reagents, pre-clinical and clinical testing, and novel chemical compounds. Since open source researchers cannot charge patients, they are unable to repay these expenses.¹¹ Second, drug patents are notoriously lucrative. Given that so many biologist-entrepreneurs have become millionaires, why should anyone donate their ideas to an open source collaboration? These are serious hurdles to collaboration. At some level, however, the arguments prove too much: Software, after all, is also expensive to make and frequently lucrative. Yet these facts have not stopped open source from colonizing and even dominating many products. Software creation is big enough to include *both* open source *and* proprietary projects. Might not similar niches exist in the complex world of drug discovery? As we shall see, closer examination reveals many locations along the drug discovery pipeline where patent-driven research and development is indeed compatible with – and may even support – open source methods.¹²

This article reviews the first tentative experiments with open source drug discovery and identifies spaces within the drug discovery pipeline where more convincing examples of open source collaboration could take root. Section II discusses the sometimes elusive definition of “open source drug discovery” and distinguishes the term from “open science” and related concepts. Section III examines recent biology collaborations that are sometimes said to resemble open source. Section IV argues that more convincing collaborations can be designed by scrutinizing the drug discovery pipeline for subtasks that match the known strengths and weaknesses of open source methods. Section V describes three relatively modest examples of open source collaborations that could plausibly be built in the next few years. Section VI discusses more ambitious, longer-term ideas for using open source methods to cut costs at the very expensive downstream end of the drug discovery pipeline. Section VII asks whether open source drug discovery needs so-called “viral” or “copyleft” licenses that prevent users from patenting improvements and, if so, whether such agreements are legal. Finally, Section VIII presents a brief conclusion.

¹⁰ Joseph A. DiMasi, Ronald W. Hansen & Henry G. Grabowski, *The Price of Innovation: New Estimates of Health Development Costs*, 22 J. HEALTH ECON. 151, 180 (2003) (averaging over failed products).

¹¹ Wesley M. Cohen, *Does Open Source Have Legs?*, in INTELLECTUAL PROPERTY RIGHTS IN FRONTIER INDUSTRIES, (Robert W. Hahn ed., 2005); Bernard Munos, *Can Open-Source R&D Reinvigorate Drug Research*, NATURE REVIEWS: DRUG DISCOVERY, Aug. 18, 2006, at 2, available at http://www.tropicaldisease.org/documents/OpenSourceDrugR&D_Nature_Munos.pdf; see also Symposium on Bioinformatics and Intellectual Property Law, Josh Lerner, *Open Source Genomics*, 8 B.U.J. SCI. & TECH. L. 254 (2002), available at <http://www.bu.edu/law/scitech/volume8/Panel6.pdf> (Open source incentives may not be sufficient “when we look at a realm like biotechnology with extremely long lags and large costs to commercialize products.”).

¹² Cf. Weber, *supra* note 8, at 217 (arguing that drug development could evolve toward a system that includes a significant “open source ecology” alongside traditional “closed source” segments).

II. WHAT DOES OPEN SOURCE DRUG DISCOVERY MEAN?

As Professor David Opderbeck shrewdly remarked, “[e]xactly what ‘open source’ means is a subject of some confusion and debate.”¹³ Strictly speaking, the phrase “open source” is limited to source code, i.e., one particularly transparent and human-readable format for recording computer code.¹⁴ To this extent, phrases like “open source biology” and “open source drug discovery” are oxymorons. That should not, of course, stop us from using them. “Words,” as Humpty Dumpty in *Alice in Wonderland* observed, “mean what I choose them to mean.”¹⁵ But we should at least reserve the label for something new. As we will see in Section III, biology collaborations often use the word “open” to describe the absence of patents. Patent-less research, however, is hardly novel. After all, scientists were already doing research for such non-monetary reasons as reputation and curiosity in Linnaeus’ time.¹⁶

If there is something new in “open source” production, then, it must be more than the traditional practices that some observers call “open science.”¹⁷ That extra something is focus, i.e., the desire to obtain not just knowledge but a specific product. Far more than in traditional science, it is this purposefulness that forces members to interact, to subordinate themselves to a larger plan, and to judge success by such suspiciously capitalist measures as consumer acceptance and market share. Properly understood, then, “open source” is less a legal category than a behavior. Practically all biology collaborations receive government or foundation funding¹⁸ and it is easy for such groups to add open source licenses to work that they would do anyway. In this case, the actual research activity – for example, how tasks are done and who participates – hardly

¹³ David W. Opderbeck, *The Penguin’s Genome, or Coase and Open Source Biotechnology*, 18 HARV. J.L. & TECH. 167, 179 (2004). See also Andres Guadamuz Gonzalez, *Open Science: Open Source Licenses in Scientific Research*, 7 N.C.J.L. & TECH. 321, 325 (2006) (“There is considerable discussion about the different definitions and variations of what is generally understood as open source software.”); STEVEN WEBER, THE SUCCESS OF OPEN SOURCE 267 (2004) (“A note of caution: As open source has begun to attract broad public attention over the last few years, the term itself has been overused as a metaphor.”).

¹⁴ The opposite of “source code” is “machine language.” Modern software is universally written (and modified) at the “source code” level. Special programs called “compilers” are then used to turn it into the efficient machine language versions that computers understand. Human programmers find it virtually impossible to understand or modify programs written in machine language. Most companies distribute compiled code to the public but keep source code versions in-house. This provides an added level of “self-help” protection over and above formal copyright sanctions.

¹⁵ LEWIS CARROLL, THROUGH THE LOOKING GLASS 33 (Signet Classic ed., 2000) (1872).

¹⁶ *Carolus Linnaeus*, WIKIPEDIA, http://en.wikipedia.org/wiki/Carolus_Linnaeus (last visited July 15, 2007).

¹⁷ See generally, Paul A. David, *Can “Open Science” be Protected from the Evolving Regime of Intellectual Property Rights Protections?*, 160 J. INSTITUTIONAL & THEORETICAL ECON. 1 (2004).

¹⁸ Arti Kaur Rai, *“Open and Collaborative” Research: A New Model for Biomedicine*, in INTELLECTUAL PROPERTY RIGHTS IN FRONTIER INDUSTRIES, *supra* note 11, at 131.

changes at all. “Open source” should mean something more than that. A useful definition should track substance, not just paperwork.

What that definition should be is not entirely clear. Even for software, definitions of “open source” tend to be arbitrary and occasionally controversial. For purposes of this article, we will paraphrase Professor Yochai Benkler’s seminal description and define open source as:

- (a) a method of producing complex economic products;
- (b) capable of supporting medium- to large-scale collaborations, potentially including thousands of people; and
- (c) organized according to signals that are neither hierarchical commands (as in firms or academic laboratories) nor prices (as in markets) but voluntary or social.¹⁹

This definition is not entirely satisfactory, since subpart (c) describes incentives negatively (i.e., *not* price, *not* hierarchical) instead of using an explicit list. This flaw is understandable for a new subject that is still being explored. More importantly, things are not as bad as they seem; as the next Section points out, social scientists have identified roughly a half dozen incentives that seem to drive all existing open source software projects. In this article, we will normally assume that viable open source drug discovery collaborations must choose from this same, relatively short list of incentives.

In what follows, we ask whether collaborations that fit this description are a feasible way to do drug discovery and if so, whether they would make society better off. We begin by surveying recent attempts to import at least some open source features into biology.

III. FIRST STEPS

The open source label is so appealing that it was only a matter of time until biology collaborations started to use it. These collaborations can usefully be grouped into five categories. The first three – software, community-wide big science projects, and databases – are fairly traditional, grant-supported activities. Because they are hierarchically organized and are not focused on producing

¹⁹ Yochai Benkler, *Coase's Penguin, or Linux and the Nature of the Firm*, 112 YALE L.J. 369, 371-75 (2002). We have expanded Benkler’s definition to include small groups who otherwise fit the definition. This reflects the fact that many open source software collaborations are small and that the same will likely be true of drug discovery. Provided that collaborations are *potentially* scaleable, *actual* size should not matter. Similarly, our third criterion expands on Benkler’s negative observation that open source collaborations differ from firms (hierarchies) and markets (price signals) by arguing that the organizing principles are primarily voluntary or social. As we will see, this seems to be a reasonable description of the open source drug discovery efforts described in Sections V and VI.

“complex economic products,” they fall outside Benkler’s criteria.²⁰ Nevertheless, their licenses are instructive and could point the way to genuine open source. The remaining two categories consist of organizations trying to organize real open source drug discovery programs. Currently, it is too early to know whether any of these initiatives will succeed.

Biology Software. Given the convergence of biology and computing, it is hardly surprising that biologists write software and that much of this code is open source.²¹ Examples include Biojava,²² BioPerl,²³ BioPython,²⁴ Bio-SPICE,²⁵ BioRuby,²⁶ Simple Molecular Mechanics for Proteins,²⁷ and Generic Software Components for Model Organism Databases (“GMOD”).²⁸ These projects use a wide variety of licenses. Some include “copyleft” elements that require users who develop improved versions to share their code free-of-charge. Examples include SMP (GNU General Public License),²⁹ Biojava (GNU Lesser GPL),³⁰ and BioPerl and GMOD (Perl Artistic License).³¹ Other licenses are more limited; these disclaim liability and require users to give appropriate author credit, but otherwise allow unrestricted commercialization. Examples include Bio-SPICE, which uses the Berkeley Software Distribution License,³² and Biopython, which offers a BSD-type license of its own devising.³³ The fact that

²⁰ See *id.* It is tempting to relax Benkler’s argument that open source is a “method for producing complex economic products” and to argue instead that these databases represent open source collaborations at the basic research stage. In this view, open source drug methods have already established a foothold in biology and it remains only to extend them downstream to later phases of the drug discovery pipeline. As with all definitions, the choice is largely arbitrary. It does, however, seem gratuitous to apply a retroactive “open science” label to academic science institutions that have existed for centuries. A relaxed definition would also jettison open source software’s remarkable focus on delivering specific products to consumers. This seems qualitatively distinct from academic databases, which – like practically all academic science – tend to grow in whichever direction(s) individual members find interesting.

²¹ Burk, *supra* note 7; see also Jonathan Dugan, *Open Source Initiatives in Bioinformatics* at 3 (Aug. 2001) (unpublished report), available at <http://smi-web.stanford.edu/auslese/smi-web/reports/SMI-2001-0902.pdf>.

²² *Bio Java Project*, WIKIPEDIA, http://biojava.org/wiki/Main_Page (last visited July 15, 2007).

²³ *BioPerl*, WIKIPEDIA, http://www.bioperl.org/wiki/Main_Page (last visited July 15, 2007).

²⁴ *BioPython*, WIKIPEDIA, http://biopython.org/wiki/Main_Page (last visited July 15, 2007).

²⁵ Bio-SPICE, <http://sourceforge.net/projects/BiOSpice> (last visited July 15, 2007).

²⁶ About BioRuby, <http://www.bioruby.org/> (last visited July 15, 2007).

²⁷ SMMP, <http://www.phy.mtu.edu/biophys/smmp.htm> (last visited July 15, 2007).

²⁸ GMOD Main Page, <http://www.gmod.org/home> (last visited July 15, 2007).

²⁹ GNU General Public License, <http://www.gnu.org/licenses/gpl.html> (last visited July 15, 2007).

³⁰ GNU Lesser General Public License, <http://www.gnu.org/licenses/lgpl.html> (last visited July 15, 2007).

³¹ Perl, The Artistic License, <http://www.perl.com/pub/a/language/misc/Artistic.html> (last visited July 15, 2007).

³² Open Source Initiative OSI- The BSD License: Licensing, <http://www.opensource.org/licenses/bsd-license.php> (last visited July 15, 2007).

³³ *BioPerl*, *supra* note 23.

most collaborations use off-the-shelf licenses suggests that the issues facing biology software are not very different from other kinds of code. This may be, as Professor Arti Rai has suggested, because software licenses have little impact on downstream drug discovery rights.³⁴

The question remains which license to choose. For non-biology software, open source software licenses are often – though not always – selected as a kind of marketing ploy to attract ideologically-motivated volunteers.³⁵ However, biology software projects are almost always publicly funded,³⁶ which leaves little or no need to attract volunteers. In this case, one might think that “he who pays the piper should call the tune,” i.e., that funding agencies should specify a single best license for all collaborations. Predictably, funding agencies have been reluctant to grasp this nettle.³⁷ The problem is that the alternative – letting individual collaborations choose – is surely misguided. At the end of the day, scientists have neither the time nor the training to choose between licenses. Furthermore, different collaborations are bound to pick inconsistent licenses. This can only complicate the legal hurdles facing follow-on projects that seek to compare and combine earlier work. Foundations would be wiser to study the problem in detail and draft uniform terms for grantees to use. Section VI will comment further on what this analysis should be.

Because bioinformatics programs are used to discover drugs, there is a definitional sense in which they must also qualify as “open source drug discovery.” At the same time, the fact that biology software – like all programs – can be produced using open source methods is hardly surprising. Clearly, it would be much more interesting if non-software projects could be organized using open source methods.

Big Science Projects. A second set of examples consists of large hierarchical team projects designed to acquire key data for an entire community. One of the earliest and most prominent examples was the SNP Consortium.³⁸ Here, Britain’s Wellcome Trust Foundation and thirteen private sector firms paid scientists to discover genome data and place them in the public domain.³⁹ Professor Dan Burk has argued that the SNP Consortium’s decision to put its discoveries in the public domain shares many features with open source.⁴⁰ However, it also made good business sense for corporate donors that wanted to block a commercial rival, Celera, from becoming “the Bill Gates of the human

³⁴ Rai, *supra* note 18, at 148.

³⁵ J. Lerner & J. Tirole, *The Scope of Open Source Licensing*, 21 J. L. ECON. & ORG. 20 (2002).

³⁶ Rai, *supra* note 18, at 140. Industry also supplies a large number of volunteers. They reportedly participate “for business reasons” or because they want to publish papers. Dugan, *supra* note 21, at 6.

³⁷ The agencies’ reticence may stem from a reluctance to endorse licenses that clash too openly with Congress’ pro-commercial philosophy as codified in the Bayh-Dole Act. See generally D. Malakoff, *Petition Seeks Public Sharing of Code*, 294 Sci. 27 (2001).

³⁸ See Maurer, *supra* note 6.

³⁹ *Id.*

⁴⁰ Burk, *supra* note 7.

genome.”⁴¹ Instead of making us cynical about open source drug discovery, the observation should give us hope. This is, after all, a case where commercial firms spent money that they had earned from patents to promote openness.

The SNP Consortium’s legacy is also visible in the Alliance for Cell Signaling (“AfCS”), a government-funded consortium of nine academic laboratories working together to map the chemical inputs and outputs that control cell behavior.⁴² AfCS members are allowed to keep data confidential to preserve publication priority. However, this embargo is temporary. Indeed, members are required to place all data in the public domain as soon as their “main findings” are accepted for publication.⁴³ Furthermore, members must also post after-acquired data “deemed relevant” to the paper’s conclusions.⁴⁴ Once posted, “all data may be used by any party for research and/or commercial purposes.”⁴⁵ Finally, AfCS members also waive patents for any discovery that results directly from Alliance funds or reagents. The stated purpose of this clause is to enhance openness between members and eliminate administrative delays related to intellectual property protection. The clause is also said to “facilitate the rapid placement of research discoveries in the public domain,” presumably by removing the temptation to delay publication in order to gain a head start on commercialization.⁴⁶

Finally, the SNP Consortium’s public domain model has received a copyleft twist from the HapMap Consortium’s \$130 million project to compare multiple human genomes to find disease-causing variations.⁴⁷ For the first two years of HapMap’s existence, users could only download data if they agreed to a “clickwrap” license promising not to file “composition of matter” patent claims for any SNP, genotype, or haplotype data based at least partly on HapMap data.⁴⁸ Although modeled on the GNU copyleft license, the HapMap license actually focused on a much narrower problem. If the group waited to distribute data until all sequencing was finished, it could deposit a complete list of variations in the

⁴¹ Maurer, *supra* note 6, at 55-56. Merck single-handedly funded a similar academic effort to put “Expressed Sequence Tags” in the public domain because it feared “upstart genomics companies . . . cornering all the rights to valuable genome information.” Hope, *supra* note 7, at 156 (quoting Roger Brent).

⁴² Alliance for Cell Signaling, *Overview of the Alliance for Cell Signaling*, 420 NATURE 703, 703-06 (2002); See also Alison Abbott, *Alliance for Cellular Signaling: Into Unknown Territory*, 420 NATURE 600, 600-01 (2002).

⁴³ Alliance for Cellular Signaling, Data Sharing Policy, <http://www.afcs.org/tools/data-sharing.html> (last visited July 15, 2007).

⁴⁴ *Id.*

⁴⁵ *Id.* AfCS’s rules probably did not impose much of a burden on members, since journals routinely require that authors make their underlying data available after publication. Andrej Sali (personal communication).

⁴⁶ Data Sharing Policy, *supra* note 45.

⁴⁷ *International Consortium Launches Genetic Variation Mapping Project*, NIH News Advisory, (Oct. 2002), available at <http://www.genome.gov/10005336> (last visited July 15, 2007).

⁴⁸ Rai, *supra* note 18, at 142-43. HapMap’s “Data Access Policy for the International HapMap Project” is no longer posted on the Internet. However, the relevant language can be found in Gonzalez, *supra* note 13, at 350.

public domain. But if it did not wait and published its data piecemeal, outside researchers could compare it against their own sequences to find and patent variations that the Consortium would eventually discover anyhow. The solution was not perfect – HapMap’s copyleft feature reportedly kept several public genome databases from using its data⁴⁹ – but it did let members publish data immediately without an AfCS-style embargo.⁵⁰ Furthermore, HapMap promised from the outset that it would waive its restrictions as more data accumulated.⁵¹ This was originally supposed to happen “around the end of calendar 2005.”⁵² HapMap lifted its restrictions in late 2004.⁵³ For now, the future of HapMap-style clickwrap restrictions is unclear. Because many journals refuse to print articles based on restricted data, future collaborations may prefer systems based on SNP Consortium-style deposits to the public domain or AfCS-style embargos. However, the jury is still out.⁵⁴

Because these projects are grant-supported and hierarchical, they fail to satisfy Professor Benkler’s requirement that genuine open source collaborations rely neither on market signals nor commands. Furthermore, they are essentially indiscriminate in their search for knowledge and, to that extent, lack open source’s focus on producing particular economic products. However, these distinctions could turn out to be details. Conceivably, their descendants may yet evolve into directions that fit Professor Benkler’s definition.

Databases. Like most sciences, modern academic biology has been constructed around a small number of large, open databases.⁵⁵ Building these resources requires extensive volunteer labor, access to unpublished information, and processes for developing communal judgments.⁵⁶ These challenges are very similar to those that a hypothetical open source collaboration, particularly for data-intensive applications based on discovering drugs *in silico*.⁵⁷

⁴⁹ *International HapMap Consortium Widens Data Access*, NIH News Release, NATIONAL HUMAN GENOME RESEARCH (Dec. 10, 2004), available at <http://www.genome.gov/12514423> (last visited July 15, 2007).

⁵⁰ *Id.*

⁵¹ *Id.*

⁵² International HapMap Project, Genotype Access Registration, <http://www.hapmap.org/cgi-perl/registration> at 3 (last visited Aug. 6, 2006).

⁵³ *International HapMap Consortium Widens Data Access*, *supra* note 49.

⁵⁴ Rebecca S. Eisenberg & Arti Kaur Rai, *Harnessing and Sharing the Benefits of State-Sponsored Research: Intellectual Property Rights and Data Sharing in California’s Stem Cell Initiative*, 21 BERKELEY TECH. L.J. 1187, at 1203-05.

⁵⁵ For a comprehensive survey of the leading biology databases, see Dugan, *supra* note 21, at 37-49.

⁵⁶ Stephen M. Maurer, Richard B. Firestone & Charles R. Scriver, *Science’s Neglected Legacy*, 405 NATURE 117 (2000); Stephen M. Maurer, *New Institutions for Doing Science: From Databases to Open Source Biology* (Nov. 19, 2003) (unpublished conference paper), available at http://www.merit.unimaas.nl/epip/papers/maurer_paper.pdf (last visited July 15, 2007).

⁵⁷ See Maurer, *New Institutions for Doing Science*, *supra* note 56.

Recent experiments have centered on finding ways for members to pool their knowledge of gene function. The most technology-intensive initiative is the Distributed Annotation System.⁵⁸ It automatically combines data from independent databases provided that the authors follow certain minimal computing conventions.⁵⁹ Other automated tools (e.g., BioCorba⁶⁰ and the Piper Project⁶¹) are designed to make peer-to-peer work flow more efficiently across geographically dispersed networks. Finally, a few proposals eschew technological solutions in favor of better social organization. Because they harvest human judgment, such low technology solutions often produce very powerful databases.⁶² One gene annotation wiki (“Wiki for Professionals”) is already in place.⁶³ Interestingly, the experiment has received approximately \$2 million in private funding.⁶⁴ Backers hope to host private versions of the system that let customers add their own proprietary data to the publicly-available pages.⁶⁵ If successful, such tools could be readily adapted to many of the open source models discussed in Section V.

Software apart, biology databases are probably the closest analogs to genuine open source drug discovery. Once again, the principal objection is that they do not focus on a specific product. However, physical science databases have long predicted the results of future experiments and it is natural to think that biology databases will one day emulate them by predicting drugs.⁶⁶ For this reason, it is reasonable to see current database collaborations as a possible stepping stone to open source drug discovery.

Cambia. Probably the best-known project with open source ambitions is a Rockefeller-funded collaboration called Cambia.⁶⁷ Its Biological Innovation for Open Society (“BiOS”) initiative covers a wide assortment of activities, many of which envisage traditional grant-supported work within Cambia itself – for example, inventing new ways to manipulate genomes, extending Cambia’s database of life sciences patents, and various leadership and advocacy

⁵⁸ BioDAS Main Page, <http://biodas.org> (last visited July 15, 2007). For a short non-technical description of BioDAS, see Dugan, *supra* note 21, at 14.

⁵⁹ *Id.*

⁶⁰ BioCORBA, <http://www.bioperl.org/wiki/BioCORBA> (last visited July 15, 2007). For a short non-technical description, see Dugan, *supra* note 21, at 15.

⁶¹ The Loci/ Piper Project Abstract, <http://www.bioinformatics.org/loci/> (last visited July 15, 2007).

⁶² On the central role of human judgment in database production, see Maurer, Firestone & Scriver, *Science’s Neglected Legacy*, *supra* note 56.

⁶³ Jim Giles, *Key Biology Databases Go Wiki*, 445 NATURE 691 (2007); Kai Wang, *Gene-function Wiki Would Let Biologists Pool Worldwide Resources*, 439 NATURE 534 (2006).

⁶⁴ Wang, *supra* note 63.

⁶⁵ Giles, *supra* note 63.

⁶⁶ Maurer, Firestone & Scriver, *supra* note 56.

⁶⁷ For a description of Cambia and its \$1 million grant from the Rockefeller Foundation, see Carina Dennis, *Biologists Launch “Open-Source Movement,”* 431 NATURE 494 (2004).

activities.⁶⁸ Here we focus more narrowly on BiOS' attempts to promote what it calls new platforms for cooperative invention, improvement and delivery of biological technologies within a dynamic "protected commons."⁶⁹ The main components of this effort so far are two draft "BiOS-compliant" Technology licenses,⁷⁰ a "BiOS-Compliant Materials Transfer Agreement" in both "Detailed" and "Simple" versions, and an "initial prototype" web site called Bioforge⁷¹ where members can work together on projects "seeded" by Cambia-developed technologies or else recommend new projects of their own.⁷² It would be nice to know whether outside volunteers have actually used Bioforge to do research. For now, the answer seems to be no.⁷³

Unless and until Bioforge starts functioning, Cambia's licensing scheme is purely theoretical. Nevertheless, the drafters have made some interesting choices. Probably the most striking is Cambia's attempt to bring every improvement that would otherwise give rise to a blocking patent within a "protected commons" where every member could use it.⁷⁴ Although Cambia argues that proprietary firms would otherwise "capture" its technology, the argument is far from self-evident. First, one would normally assume that placing technology in the public domain is already enough to protect it from "capture."

⁶⁸ Cambia, The Cambia BIOS Initiative at 3, <http://www.BiOS.net/daisy/BiOS/10/version/live/part/4/data> (last visited July 15, 2007).

⁶⁹ *Id.*

⁷⁰ Cambia, BiOS-Compliant Licensing Listing, <http://www.cambia.org.au/daisy/bios/licenses/398/2531.html> (last visited July 15, 2007).

⁷¹ Cambia, Bioforge: An Online Community for Biotechnology Innovation, <http://www.bioforge.net/forge/index.jspa> (last visited July 15, 2007).

⁷² Cambia, Forum, Project Recommendations, <http://www.bioforge.net/forge/forum.jspa?forumID=26> (last visited July 15, 2007). In the software world, open source collaborations seldom if ever produce code from scratch. Instead, organizers almost always show would-be members working (if frequently incomplete or flawed) software to prove that the proposed project is feasible, interesting, and potentially useful. Josh Lerner & Jacques Tirole, *Some Simple Economics of Open Source*, 52 J. INDUS. ECON. 197, 197-232 (2002).

⁷³ Since Bioforge sites are closed to non-members, it is hard to know whether any research is taking place. It is reasonably clear that some companies have signed licenses. Cambia, About BiOS (Biological Open Source) Licenses, <http://www.bios.net/daisy/bios/licenses/398.html> (last visited July 15, 2007). *But see Open Source Biotech: Borrowing the Software Model to Spur Innovation in Life Sciences Finds Common Ground with IT*, 3 RED HERRING 14 (April 17, 2006) (Cambia has so far failed to attract very large players like "a Monsanto or a . . . Merck."). However, the author has found no evidence that Cambia has taken the next step of organizing licensees to improve its technology. To the contrary, Cambia's tentative description of BioForge suggests that such activities remain hypothetical. Cambia, *supra* note 73. This impression is reinforced by Bioforge's projects pages, none of which contain more than a dozen or so short postings. BioForge, Projects: All Projects, <http://www.bioforge.net/forge/kbcategory.jspa?categoryID=1> (last visited July 15, 2007).

⁷⁴ *See, e.g.*, The CAMBIA BiOS License for Plant Enabling Technology Version 1.3.1 at ¶ 1.6, <http://www.bios.net/daisy/PELicense/751/387.html> (last visited July 15, 2007) (defining "improvements" subject to grantback as "any improvement . . . which . . . but for the terms of this License Agreement cannot be used without infringing a valid claim in an unexpired [Cambia] Patent.").

In the software world, this answer is not always sufficient because code needs to be maintained and will wither away if it is not used.⁷⁵ This objection has much less force for biological inventions, however, whose functionality is permanently encoded into the living world. A second, more sophisticated argument echoes the HapMap rationale: At least in principle, companies could monitor the collaboration's work and then rush in to patent trivial improvements that would have been discovered in any case.⁷⁶ However, in this case the HapMap solution – a temporary embargo – ought to be sufficient. Finally, a third and more radical argument assumes that most (though presumably not all) inventors who apply for patents would still make and publish discoveries if intellectual property incentives did not exist. In this case, restrictions on patenting could significantly increase the fraction of public domain inventions while only slightly reducing the total volume of innovation.⁷⁷ We return to this tradeoff in Section VI below. For now, we remark only that many open source software collaborations let members make proprietary improvements⁷⁸ and that Cambia's decision to write viral terms makes its licenses very complex.⁷⁹

Finally, there is a real question whether Cambia's vision should be called "open source" at all. Contractually, the scheme is more or less identical to a commercial patent pool in which licensees receive the non-exclusive right to use a patented technology in exchange for promising to grantback patented improvements to other pool members.⁸⁰ The difference, according to Cambia, is that its protected commons would be "accessible to all."⁸¹ This would indeed be true if Cambia charged no royalty. However, Cambia actually proposes to charge commons members an annual "subscription fee" ranging from \$10,000 to

⁷⁵ Stephen M. Maurer & Suzanne Scotchmer, *Open Source Software: The New Intellectual Property Paradigm*, in HANDBOOKS IN INFORMATION SYSTEMS, VOL. 1 (T. Hendershott ed., 2006).

⁷⁶ One might think that trivial improvements would fail the patent law's non-obviousness requirement. 35 U.S.C. §§ 102-103 (2000). This, however, assumes that the non-obviousness requirement is well-defined and/or that the US Patent and Trademark Office enforces it. Both assumptions are debatable.

⁷⁷ Alfonso Gambardella & Bronwyn Hall, *Proprietary vs. Public Domain Licensing of Software and Research Projects* 13-15 (Nat'l Bureau of Econ. Research, Working Paper No. 11120, 2005).

⁷⁸ See, e.g., *Berkeley Software Distribution License*, WIKIPEDIA, http://en.wikipedia.org/wiki/BSD_license (last visited July 16, 2007)..

⁷⁹ Because Cambia hopes to attract commercial support, it allows users to patent products made with its tools. However, Cambia concedes that the distinction between tools and product is "a moving target" that will likely be "difficult and contentious." Cambia, The Cambia BIOS Initiative, <http://www.BiOS.net/daisy/BiOS/10/version/live/part/4/data> (last visited July 15, 2007).

⁸⁰ See, e.g., *Transparent-Wrap Mach. Corp. v. Stokes & Smith Co.*, 329 U.S. 637 (1947). Cambia's licenses also recognize licensees' right to keep secrets from each other. See, e.g., Cambia, License for Plant Enabling Technology Version 1.3.1, *supra* note 74, at ¶ 1.6 (licensing obligations do not extend to improvements that are "actively maintained as a proprietary trade secret for use in BiOS Licensee's business . . ."). This approach seems far removed from stereotypical open source models in which advances come from distributing improvements so that "many eyeballs" can critique them.

⁸¹ Cambia, The Cambia BIOS Initiative, <http://www.BiOS.net/daisy/BiOS/10/version/live/part/4/data> (last visited July 15, 2007).

\$150,000.⁸² Even for biotechnology companies, this figure is substantial.⁸³ As in a conventional commercial patent pool, profit-maximizing businesses would not join the commons unless they expected membership to generate sufficient earnings to cover their royalty fees. Since these earnings would be inversely proportional to the number of companies already using the technology, we expect the commons to stop growing at roughly the same size that commercial patent pools do.⁸⁴ While the arrangement might still benefit society,⁸⁵ it would not be “accessible to all” in any meaningful sense. Moreover, protected commons membership would be further restricted by Cambia’s internal procedures, which provide that new members must be “approved by a project curator” according to “a member approval process.”⁸⁶ Project curators drawn from industry would have an obvious incentive to prevent competitors from joining the commons.⁸⁷

For now, it is hard to say more than this. Cambia’s draft licenses and BioForge’s website have laid the groundwork for an elaborate experiment. Licensees may accept openness, they may insist on patent pool-type membership

⁸² Cambia, License for Plant Enabling Technology Version 1.3.1, *supra* note 74, at ¶¶ 6.1, 6.2 & Annex D. Cambia has also said that it hopes companies will agree to contribute in-kind research and development resources to the commons. Cambia, A Hypothetical Scenario: A Future Researcher Uses BioForge, <http://www.bioforge.net/forge/entry.jsps?externalID=33> (last visited July 16, 2007) (“If the project is particularly exciting, or a case can be made for its importance, one or more of the many public and private institutional investors and funders – foundations, governments, businesses – who also access the BioForge . . . can establish a revenue base from which to draw to overcome resource limitations on key players.”). Such contributions would constitute royalties in their own right.

⁸³ By way of comparison, companies paid a \$10,000 annual royalty during the 1980s for access to the very basic Cohen-Boyer methods for combining and transplanting genes. While licensors claimed that they had deliberately set royalties low to promote access, the fee was nevertheless a “genuine burden for many small biotech startups.” ARTHUR KORNBERG, *THE GOLDEN HELIX: INSIDE BIOTECH VENTURES* 241 (1995). The Cohen-Boyer fee is about \$18,000 in current dollars.

⁸⁴ See, e.g., Stephen M. Maurer & Suzanne Scotchmer, *Profit Neutrality in Licensing*, 8 AM. L. & ECON. REV. 476, 485-86 (2006). The effect would be largest where companies competed directly with one another.

⁸⁵ Whether or not it deserves to be called “open source,” a Cambia-led patent pool could be good public policy. An effective pool would provide competition for Monsanto, turning that company’s current monopoly of techniques for manipulating the plant genome into a duopoly. If Cambia’s pool embraced companies that competed against each other, prices would fall still further.

⁸⁶ Cambia, The Cambia BIOS Initiative, *supra* note 81. Furthermore, not all members are equal. Instead, curators have the power to define custom roles that grant specific permissions to users in their project and users are only able to perform the actions that are permissible under that role. Therefore, it is possible for some users to only have read-only access to a project while others can commit changes to the version control repository. *Id.* A second Cambia document suggests that a BioForge Portfolio manager would “assist[] in the formation of the group, and the solicitation of interested individuals or institutions to its critical mass.” Cambia, A Hypothetical Scenario: A Future Researcher Uses BioForge, <http://www.bioforge.net/forge/entry.jsps?externalID=33> (last visited July 16, 2007).

⁸⁷ Curators’ ability to exclude would presumably be limited by the antitrust laws. This, however, is not a high standard. *Transparent-Wrap*, 329 U.S. at 637.

restrictions, or they may refuse to participate on any terms at all. Only time will tell.

Host Sites. The software community maintains several on-line sites where new volunteers can find, join, and work on open source projects. The largest of these, Sourceforge, currently hosts nearly 150,000 software projects,⁸⁸ although only a small fraction of these are active.⁸⁹ Not surprisingly, there are at least two analogous sites dedicated to hosting open source biology and drug discovery software. The Open Bioinformatics Foundation seeks to enhance standard developer toolkits so that they can be used more effectively in biology.⁹⁰ Its work is tightly focused on a handful of programs including BioPerl, BioPython, BioJava, BioXML, BioCorba, and BioDAS.⁹¹ A second organization, Bioinformatics.org,⁹² is more agnostic. Like Sourceforge, it offers an on-line home to any open biology software collaboration that needs one. It currently claims 18,000 members and hosts over 250 projects.⁹³ Significantly, both collaborations build on tightly knit physical contacts among scientists and institutions. These include an annual Bioinformatics Open Source Conference⁹⁴ that attracts hundreds of participants.

There are also several sites dedicated to promoting open source biology collaborations beyond software. We have already mentioned Cambia's Bioforge site.⁹⁵ Like the Open Bioinformatics Foundation, it tends to focus on a comparably small number of Cambia-initiated technologies and specifies what licenses new projects must use. A second site, operated by The Synaptic Leap ("TSL"),⁹⁶ lets volunteers choose their own projects and licenses. TSL currently has on-line collaboration pages dedicated to malaria, schistosomiasis, toxoplasmosis, and tuberculosis.⁹⁷ Each is hosted by a different volunteer-expert, although none is currently active.⁹⁸ In addition to providing a focus for new projects, TSL is developing collaboration software that goes beyond simple e-mail threads and wiki-pages. The current plan is to test the system on volunteers

⁸⁸ Source Forge, Inc. Home Page, <http://sourceforge.net/> (last visited July 16, 2007).

⁸⁹ Stefano Comino, Fabio Manenti & Maria Laura Parisi, *From Planning to Mature: On the Determinants of Open Source Take-Off*, (July 2005) (unpublished paper), available at http://opensource.mit.edu/papers/Comino_Manenti_Parisi.pdf).

⁹⁰ OBF Home Page, <http://open-bio.org> (last visited July 16, 2007).

⁹¹ Short descriptions of these projects can be found in Dugan, *supra* note 21.

⁹² Bioinformatics Organization, <http://Bioinformatics.org/about/> (last visited July 16, 2007). See also Bioinformatics.Org, Bioinformatics.Org: 2002 Organization Plan, <http://bioinformatics.org/about/plan-20020920.pdf> (last visited July 16, 2007).

⁹³ *Id.* Like Sourceforge itself, many projects are empty and inactive. See Bioinformatics.org, *Full Project Listing*, http://bioinformatics.org/softwaremap/?form_cat=3 (last visited July 16, 2007).

⁹⁴ See, e.g., BOSC 2006, http://open-bio.org/wiki/BOSC_2006 (last visited July 16, 2007). A short description of the second BOSC conference can be found in Dugan, *supra* note 21.

⁹⁵ Bioforge, <http://www.bioforge.net/forgel/index.jspa> (last visited July 16, 2007).

⁹⁶ The Synaptic Leap, <http://thesynapticleap.org/?q=welcome> (last visited July 16, 2007).

⁹⁷ *Id.*

⁹⁸ *Id.*

doing malaria research. If the collaboration produces solid science, minimally-modified versions of TSL's code can then be adapted to other collaborations and diseases. Finally, at least one commercial venture has created tools for open source sharing. Collaborative Drug Discovery, Inc. ("CDD") offers data management software that lets users choose whether their biology data is "100% confidential," shared with a finite list of approved recipients, or open to the world.⁹⁹ CDD clearly sees the open source drug discovery idea as a market for its products.¹⁰⁰

Like Sourceforge, these sites are not open source collaborations per se, but only facilitators. We still do not know whether open source drug discovery initiatives can attract volunteers or, if they do, whether they can make useful discoveries. Organizations like Bioforge, TSL, and CDD do, however, increase the chances of success.

IV. DESIGN PRINCIPLES

The existing biology collaborations either fail to fit our definition of open source or are too preliminary to evaluate. While this observation is discouraging, it actually says very little about whether open source drug discovery is possible. The conventional arguments against open source drug discovery are (a) that open source has no way to recover its costs from patients, and (b) that patent incentives will drive out open source voluntarism wherever they meet. However, these claims are hardly airtight. First, patents and patient revenues are more or less irrelevant to a large class of pharmaceutical research and development problems. These include diseases of the developing world,¹⁰¹ bioweapons,¹⁰² and so-called orphan diseases.¹⁰³ Here, progress can *only* be made if government and non-profit programs cover researchers' costs. These initiatives could well include a role for open source. Second, commercial drug discovery is not monolithic. Instead, it consists of roughly a dozen innovation steps, each of which requires its own specialized personnel, equipment, and skill sets.¹⁰⁴ In the software world, companies often compete at one level (e.g., applications programs) while supporting open source cooperation at another (e.g., operating

⁹⁹ Collaborative Drug Discovery, Inc., http://www.collaborativedrug.com/pages/product_info (last visited July 16, 2007).

¹⁰⁰ Posting of Barry Bunin to <http://thesynapticleap.org/?q=node/140> (Jan. 7, 2007, 09:51 EST).

¹⁰¹ Of the 1,233 new drugs ("new chemical entities") licensed between 1975 and 1996, only thirteen treated neglected diseases. O. Trouiller & P.L. Olliaro, *Drug Development Output from 1975 to 1996: What Proportion for Tropical Diseases*, 3 INT'L J. INFECTIOUS DISEASE, 61, 61-63 (1999).

¹⁰² *Creating a Biodefense Industry: BioShield II: Hearing Before the S. Judiciary and HELP Comm.*, 108th Cong. (2004) (statement of Sen. Joseph Lieberman), available at <http://lieberman.senate.gov/documents/reports/BiOShieldtestimony0604.pdf>.

¹⁰³ Chris Hartman-Siantar (personal communication).

¹⁰⁴ Stephen M. Maurer, *Choosing the Right Incentive Strategy for Research and Development in Neglected Diseases*, 84 WORLD HEALTH ORGANIZATION BULLETIN 376, 376 (2006).

systems).¹⁰⁵ Therapeutics companies may similarly find it in their dollars-and-cents interest to support open source drug discovery at various points along the drug discovery pipeline.¹⁰⁶

The challenge now is to find places in the drug discovery pipeline where recognizably open source methods could actually flourish. The fact that such niches have not been colonized already suggests that they are subtle. For this reason, simple analogies to computing will not provide much guidance. This article argues that a more systematic approach is needed. There is now a large and insightful literature explaining when open source software production works and – just as importantly – when it offers more social benefits than alternative methods like patents.¹⁰⁷ One way to exploit these insights is to scrutinize the long and complex drug discovery process for niche activities where open source collaborations would be simultaneously feasible and useful. This article begins the process by identifying and discussing five specific examples.

The starting point for this analysis is to ask where open source incentives are likely to be effective. Table 1 describes typical incentives that are known to drive open source collaborations in software production. A few of these (e.g., production for the inventor's own use or production to sell a related good or service) seem to lack clear analogs in drug discovery. However, many other incentives do apply. For example, it is not hard to imagine that biologists would join collaborations to learn new skills (education), demonstrate those skills to others (signaling), or donate their services from a sense of altruism.¹⁰⁸ For drugs with commercial potential, employers could also pay employees to “volunteer” for cold-blooded business reasons. For example, a firm could decide that open source collaborations were a good way to share costs with companies that do not produce competing products.¹⁰⁹ Or it could decide that an open product would make its own proprietary inventions more valuable.¹¹⁰ Sections V and VI identify five different locations along the drug discovery pipeline where the foregoing incentives could be strong enough to drive an open source collaboration.

Of course, arguing that an open source collaboration is possible is not the end of the story. We should also ask whether a successful open source collaboration would be good for society. In general, open source methods have both strengths and weaknesses compared to patents. Consumers benefit chiefly by (a) low prices, which facilitate the widest possible consumption, and (b) increased transparency, which makes it easier to judge the quality of both existing and proposed products.¹¹¹ Similar benefits are likely to be important at various points along the drug discovery pipeline. Open source software also has

¹⁰⁵ Stephen M. Maurer & Suzanne Scotchmer, *Open Source Software: The New Intellectual Property Paradigm*, in HANDBOOKS IN INFORMATION SYSTEMS, VOL. 1 (T. Hendershott ed., 2006).

¹⁰⁶ Maurer, *Choosing the Right Incentive Strategy*, *supra* note 104.

¹⁰⁷ For a review of this literature, see Maurer & Scotchmer, *supra* note 105.

¹⁰⁸ *Id.* at 287-300.

¹⁰⁹ *Id.* at 288-89.

¹¹⁰ *Id.* at 290-92.

¹¹¹ *Id.* at 306-08.

disadvantages, most notably in its tendency to decouple product creation from the price signal and, hence, consumers' needs. Ironically, this is not always a disadvantage for drug discovery. Indeed, efforts by western governments and foundations to find cures for neglected diseases normally assume that market-driven research and development signals are hopelessly inadequate and should be overridden. Section VII argues that open source collaborations would be welfare improving for each of the five niches described below.

Finally, open source methods often shift research and development costs from users onto innovators. This raises potential fairness issues. These issues probably do not matter much for neglected diseases, where it is probably acceptable to shift research and development costs from developing world patients (who cannot pay in any case) to volunteer researchers. The question of fairness is not as clear for rich nation diseases. However, in this case volunteers often receive corporate support. These costs will almost always be passed back to patients through the patent system.

V. NEXT STEPS

Here, we describe three open source drug discovery collaborations that are either currently being organized or could be organized within the next few years. Success would potentially lay the foundation for some of the more ambitious projects described in Section VI.

In Silico Drug Discovery. During the 1980s, biologists began using computers and large databases to do biology research “*in silico*.”¹¹² The basic idea was to identify proteins encoded by the genome that cause disease (“drug targets”) by noticing subtle patterns in their sequence and/or structure, and then design novel molecules (“drug candidates”) to turn off the proteins’ functions.¹¹³ The main challenge for such projects – motivating large numbers of skilled workers to find and patch obscure flaws – is strongly reminiscent of Linus Torvalds’ dictum that “given enough eyeballs, all bugs are shallow.”¹¹⁴ Given that the human genome is not much larger than Microsoft Windows,¹¹⁵ it is

¹¹² Andrej Sali (personal communication). “*In silico*” is a Latin expression used to mean “performed on computer or via computer simulation.” *In silico*, WIKIPEDIA, http://en.wikipedia.org/wiki/In_silico (last visited Jan. 29, 2008).

¹¹³ Andrej Sali (personal communication).

¹¹⁴ ERIC RAYMOND, *THE CATHEDRAL AND THE BAZAAR* 30 (1999) (quoting Linus Torvalds).

¹¹⁵ Current versions of Windows contain roughly 2.5 billion bits of information. See, e.g., Windows Vienna – Two and a Half Years From Now, <http://www.windowsvienna.com> (last visited July 15, 2007). By comparison, the human genome contains about three billion base pairs. See *Human Genome*, WIKIPEDIA, http://en.wikipedia.org/wiki/Human_genome (last visited July 16, 2007). Admittedly, readers should take this observation with a large grain of salt. The genomic “alphabet” uses four characters and therefore contains more information than the binary language of software. More importantly, users trying to decipher the genome have no source code to consult.

reasonable to think that methods which work in computing might also work for drug discovery.¹¹⁶

Since 2004, I have worked with Professors Arti Rai and Andrej Sali¹¹⁷ to establish a Tropical Disease Initiative ("TDI")¹¹⁸ where volunteers could comb on-line databases and run sophisticated computational chemistry simulations to develop new drug ideas. Just as most open source software collaborations start with a preexisting (if limited and buggy) code base, TDI will start with a kernel of possible targets supplied by a core group of researchers. Volunteers will then be organized around three separate tasks. First, they will search electronic databases for additional lines of evidence that the target actually does or does not control disease.¹¹⁹ The resulting information will be recorded separately for each target in an on-line document ("gene card") that members can continually update. Second, members will use sophisticated computational chemistry software to predict new chemical compounds ("drug leads") that (a) bind to proposed targets, and (b) possess certain characteristics needed to make a working drug. Compounds predicted by more than one simulation are likely to be especially promising. Finally, TDI members will reach a collective judgment identifying the most promising drug leads for further funding.

Like all open source drug discovery collaborations, TDI is an experiment. In scientific terms, skeptics point to the fact that commercial programs have found very few drugs using computers.¹²⁰ However, it is reasonable to think that TDI can do better, chiefly because developing world diseases have been studied far less than rich nation diseases in which all the easy drugs have been known for decades. In social terms, TDI skeptics ask whether the initiative can attract volunteers. TDI argues that there are many reasons for members to volunteer including ideology, education (i.e., the chance to learn computational biology), and signaling (i.e., the chance to advertise skills to potential employers). These incentives are similar to the ones that drive many open source software

¹¹⁶ See generally Stephen M. Maurer, Richard B. Firestone & Charles R. Scriver, *Science's Neglected Legacy*, 405 NATURE 117 (2000). The fifty year-old tradition of large physics databases provides further grounds for thinking that *in silico* biology can be organized along open source lines. Because physics variables are related to each other through "conservation laws," knowing, say, three variables often lets researchers calculate the fourth. Editors of large physics databases routinely exploit this fact to predict values for variables that have never been measured. Using biology databases to predict drugs would be a similar, if decidedly more focused and goal-oriented activity. There is also a second similarity. While big physics databases are usually built around small paid staffs, most of the information is collected and compiled by volunteer editors. At least in physics, incentives like curiosity and reputation seem to be quite powerful. *Id.*

¹¹⁷ Stephen M. Maurer, Arti Kaur Rai & Andrej Sali, *Finding Cures for Tropical Diseases: Is Open Source an Answer?*, 1 PUB. LIBR. SCI.: MED. 56 (2006).

¹¹⁸ Tropical Disease Initiative, <http://www.tropicaldisease.org/?page=home> (last visited July 16, 2007).

¹¹⁹ Members will also be asked to add new potential targets to the original kernel.

¹²⁰ Iaian M. Cockburn, *Is the Pharmaceutical Industry in a Productivity Crisis?* 7 INNOVATION POL'Y & ECON. 1 (2007) (describing recent perceived failures in commercial drug discovery); Brian K. Shoichet, *Virtual Screening of Chemical Libraries*, 432 NATURE 862, 862 (2004) (promise of computational methods for finding ligands is "still largely unfulfilled").

collaborations¹²¹ and should be particularly attractive to academics and graduate students. TDI would also allow students and scientists in the developing world to trade their passion and detailed local knowledge for the chance to work with and learn from leading rich nation research groups. Such interactions may never occur outside TDI's online world.

Chemistry. Despite years of effort, biologists' ability to predict successful drug candidates *in silico* remains limited. For this reason, it is important to supplement computerized research with physical experiments. The difference, of course, is that physical experiments are expensive. This means that open source collaborations must simulate a world in which reagents – like computer time – cost nothing. To some extent, this world already exists since academic labs seldom bother to meter the costs of small-scale experiments. Furthermore, academic grantees can sometimes scrounge chemicals from other grants.¹²² However, one suspects that funding agencies would have only limited tolerance for such behavior. The challenge is to find useful research and development projects that can be done within these constraints. Professor Matthew Todd believes that one such problem involves finding a better way to synthesize praziquantel, today's preferred drug for treating schistosomiasis.¹²³ Current methods indiscriminately make two mirror-image versions of the molecule, only one of which is biologically useful. Producing a pure version of the active ingredient would make it easier to administer the drug to children and deliver larger doses to adults.¹²⁴ "There are two ways in which organic chemists can help. One is to design synthetic routes or share experience on those being discussed. The other is to attempt one or two steps and share results, so that an optimized route may be arrived at."¹²⁵ Significantly, only the second step requires physical resources. The first is a matter of cleverness and imagination.

Stem Cell Lines. In the software world, the price that an operating system commands often has very little to do with its inherent quality or cleverness. Instead, consumers may purchase it because other people already own it or have written applications for it or for other basically historical reasons. Because of these network effects,¹²⁶ even an indifferent product can end up commanding high prices. This situation is particularly frustrating for free-lance programmers who customize software for individual customers since it means splitting their wages with a monopolist. During the 1990s, many of these professionals famously created LINUX as an alternative to Windows. For them, working on

¹²¹ Maurer & Scotchmer, *supra* note 105.

¹²² Maurer, Rai & Sali, *Finding Cures, supra* note 117; Yochai Benkler, *Commons-Based Strategies and the Problems of Patents*, 305 SCIENCE 1110 (2004).

¹²³ Matthew Todd, *Enantioselective Synthesis of Praziquantel*, <http://thesynapticleap.org/?q=schist/projects>.

¹²⁴ *Id.*; Thomas B. Kepler *et al.*, *Open Source Research – The Power of Us*, 59 AUST. J. CHEM. 291 (2006).

¹²⁵ Todd, *supra* note 123.

¹²⁶ SCOTCHMER, *supra* note 1, at 289–318.

LINUX meant “creating code that they will never have to pay someone to use again.”¹²⁷

For biologists, the closest analog to an operating system is a cell line. In the case of stem cells, therapeutics companies are already trying to coax individual cell lines to make heart muscle and other replacement tissues for the human body. In theory, each company could use its own lines. That, however, would mean forfeiting the chance to learn from past experiments. In practice, biologists prefer to adopt lines that have been used before (offer “more experience”) even if they are not otherwise known to be “better.”¹²⁸ As with Windows, this preference suggests a tipping dynamic in which initially popular lines become steadily more popular over time. Evidence for this phenomenon can be found in stem cell requests processed by The National Stem Cell Bank. One-third of all requests specify just one of the thirteen available lines and the top three lines account for almost two-thirds (sixty-two percent) of all requests.¹²⁹

It is worth asking what these dynamics would mean to a therapeutics company trying to develop stem cell-based treatments. In order to invent products, the company would already need a patent license from the Wisconsin Alumni Research Foundation (“WARF”). But what if tipping creates a second monopolist in cell lines? Some observers have suggested that challengers may be able to invent around¹³⁰ or invalidate¹³¹ WARF’s patents. Therapeutics companies will gain nothing from this, however, if tipping effects create a cell line monopolist in the meantime. The situation will be even worse if WARF’s patents are upheld. There is an old result in economics that if two monopolists own two (or more) resources needed to make a product the total price will be higher than if a single monopolist owned the inputs.¹³² All of this suggests that therapeutics companies have a strong interest in encouraging employees to develop open cell lines.

The timing is excellent. President Bush’s freeze on federal funded research on cell lines created after 2000 means that practically all incumbent lines have

¹²⁷ Siobhán O’Mahoney, *Guarding the Commons: How Community-Managed Projects Protect Their Work*, 32 RES. POL’Y 1179, 1195 (2003).

¹²⁸ National Stem Cell Bank, Frequently Asked Questions for Requesting Stem Cells, http://www.wicell.org/index.php?option=com_content&task=blogcategory&id=124&Itemid=190 (last visited July 16, 2007).

¹²⁹ National Stem Cell Bank, HIH Registered Calls, http://www.wicell.org/index.php?option=com_oscommerce&Itemid=130 (last visited July 16, 2007).

¹³⁰ Kenneth Taymor, Christopher Scott & Henry T. Greely, *The Paths Around Stem Cell Intellectual Property*, 24 NATURE BIOTECH. 411, 411-13 (2006).

¹³¹ Christopher D. Hazuka, *Supporting the Work of Lesser Geniuses: An Argument for Removing Obstructions to Human Embryonic Stem Cell Research*, 57 MIAMI L. REV. 157 (2002). The U.S. Patent and Trademark Office (“PTO”) invalidated three of WARF’s patents in April 2007. WARF has appealed the PTO’s action. Karen Kaplan, *Patents for Human Stem Cells Invalidated*, L.A. TIMES, Apr. 3, 2007, at A18.

¹³² ANTOINE AUGUSTIN COURNOT, RESEARCHES INTO THE MATHEMATICAL PRINCIPLES OF THE THEORY OF WEALTH (Nathaniel T. Bacon trans., Macmillan 1897) (1838).

been damaged by aging and contamination.¹³³ If and when the freeze is lifted, observers expect a flood of new cell lines. At this point, today's incumbents will disappear and the tipping race will start over. Of course, it is possible to imagine companies donating open lines to the public domain while stopping short of a full open source collaboration. However, open lines would not mean much unless there were also open methods to cultivate and exploit them. Recent game theory research confirms the intuition that companies that produce non-competing therapies should frequently find it in their interest to share information.¹³⁴ Indeed, even companies that compete directly can still profit from early-stage research and development sharing.¹³⁵ These results do not guarantee that open source sharing will emerge, but they do make it more likely.

VI. DISTANT, AMBITIOUS GOALS

So far, we have discussed pre-clinical projects. However, seventy-five percent of the cost of new drugs takes place after clinical trials begin.¹³⁶ For open source to be truly transformative, it must colonize this space. Using open source to organize clinical trials would be far more ambitious than the ideas discussed in Section V and could easily take a decade or more to achieve. The question remains whether such collaborations make sense. We start with the simplest case in which an open source collaboration would conduct post-approval ("Phase IV") tests to expand FDA approval for an existing drug. We then ask whether drug companies have any reason to support open source participation at the more expensive Phase I, II and III stages of testing.

Phase IV Trials. At first blush, the idea of downstream open source seems counterintuitive. Who will pay for the drugs and physicians? However, as Professors DeMonaco, Ali, and Von Hippel have pointed out, there is at least one context in which the drugs and physicians are already paid for.¹³⁷ Physicians often discover that drugs have so-called "off label" uses not previously considered by the FDA. At this point, manufacturers typically conduct additional

¹³³ Scott LaFee, *Before Stem Cells Get Used in Therapy There's the Science of Making Them*, SAN DIEGO UNION TRIBUNE, Feb. 9, 2005, at F-1 (existing cell lines suffer from aging, accumulated mutation, and genetic material from mouse "feeder cells" substrate).

¹³⁴ See, e.g., Dietmar Harhoff, Joachim Henkel & Eric von Hippel, *Profiting from Voluntary Information Spillovers: How Users Benefit By Freely Revealing Their Innovations*, 32 RES. POL'Y 1753 (2003).

¹³⁵ Nisvan Erkal & Deborah Minehart, *Optimal Sharing Strategies in Dynamic Games of Research and Development* (unpublished paper), available at <http://www.cireq.umontreal.ca/activites/papiers/05-06minehart.pdf> (last visited July 16, 2007).

¹³⁶ See, e.g., Global Alliance for TB Drug Development, *Economics of TB Drug Development* 38 (2001), available at [http://66.216.124.114/pdf/Economics%20Report%20Full%20\(final\).pdf](http://66.216.124.114/pdf/Economics%20Report%20Full%20(final).pdf) (last visited July 16, 2007).

¹³⁷ Harold J. DeMonaco, Ayfer Ali & Eric Von Hippel, *The Major Role of Clinicians in the Discovery of Off-Label Drug Therapies*, (M.I.T. Sloan, Working Paper No. 4552-05, 2005), available at <http://ssrn.com/abstract=780544>.

tests (“Phase IV trials”) to persuade the FDA to amend and expand its original approval. Given that the healthcare system already pays for treating physicians and drugs, a Phase IV open source project need only persuade members to take, report, and analyze data. Here, motives like advancing medicine, reputation, and showing off skills to potential employers may be sufficient. Professors. DeMonaco, Ali, and Von Hippel argue that the current system for finding “off label” uses suffers from “major inefficiencies” and that a model similar to “open source software” could improve matters.¹³⁸

Naively, an open source Phase IV collaboration would put its discoveries in the public domain. In this case, open source would play the same role that it does in software: reducing product prices so that more consumers can afford them. Alternatively, aggressive drug companies might think of a way to patent the new uses. Thinking of such tactics, however, is not the same as using them. Even if drug companies never obtained a patent, they would still benefit from increased sales. Waiving patent claims to these new uses might be an acceptable trade if it persuaded more open source volunteers to join the project.

Phase I - III Trials. A successful Phase IV open source collaboration would inevitably encourage volunteers to organize pre-clinical trials as well. Since the healthcare system does not normally pay for Phase I – III trials, Big Pharma would have to support these projects. Interestingly, it might have a business reason to do so. In the late 1990s, drug companies spent \$802 million for each product that they delivered to market.¹³⁹ A large fraction of these costs was devoted not so much to discovering drugs as persuading the FDA that Big Pharma’s claims were correct. Unfortunately, regulators are right to be skeptical. Over the years, there have been repeated scandals in which researchers and clinicians have falsified data to keep lucrative research and development contracts alive.¹⁴⁰ In principle, the FDA could relax its paperwork requirements – and drug companies could reduce their testing costs – if the process was more trustworthy.

Section V argued that open source is a natural vehicle for achieving transparency and/or insulating data from commercial bias. Instead of conducting their own trials, drug companies could simply make new compounds available at prices comparable to treatments already on the market. Open source volunteers would then collect and analyze data for such traditional reasons as reputation and a desire to help patients. Like today’s commercial software companies,¹⁴¹ Big Pharma could further encourage volunteers by providing free support tools, hosting conventions, and other benefits. The main challenge for such a scheme

¹³⁸ *Id.*; See also Weber, *supra* note 8, at 268-69 (asking whether physicians could use open source methods to trade knowledge about rare medical conditions).

¹³⁹ Joseph A. DiMasi, Ronald W. Hansen & Henry G. Grabowski, *The Price of Innovation: New Estimates of Health Development Costs*, 22 J. HEALTH ECON. 151, 180 (2003) (averaging over failed products).

¹⁴⁰ See, e.g., ALEXANDER KOHN, FALSE PROPHETS 176-80 (1986) (collecting instances of fraud by commercial laboratories and physicians).

¹⁴¹ Maurer & Scotchmer, *Open Source Software*, *supra* note 105, at 301.

would be persuading volunteers to test a patented drug.¹⁴² Fortunately, there is limited data from the software world suggesting that commercial rights can coexist with voluntarism.¹⁴³ Big Pharma could also attract volunteers by promising to donate its patents to the public domain after a fixed period of years.

VII. WILL OPEN SOURCE DRUG DISCOVERY MAKE SOCIETY BETTER OFF?

Commentators sometimes forget that open source drug discovery is not an end in itself.¹⁴⁴ Instead, it is only worth doing if it benefits society. In practice, these benefits tend to be different for each of the projects described above.

Neglected disease collaborations (e.g., TDI) would offer three types of benefits. First, volunteers matter. By almost any standard, research and development for neglected diseases remains badly underfunded. Open source lets volunteers pitch in by aggregating small “granular” contributions of labor and expertise into a significant research and development effort.¹⁴⁵ These contributions are especially useful because they happen to focus on the very earliest stages of the drug discovery pipeline that are most in danger of running dry.¹⁴⁶ Second, open source offers significant savings. Because open source drugs are by definition generic, western governments and foundations would pay nothing for the right to develop and (eventually) manufacture them. Open source drug candidates would also provide competition for patented drug ideas, driving down the royalties that owners could otherwise charge to sponsors. Finally, open source would make drug discovery more transparent. In conventional commercial research, Big Pharma companies know that biotech companies have both the opportunity and motive to suppress adverse results. For this reason, they seldom invest in new ideas unless and until those ideas have reached the animal testing stage. For neglected disease research, this presents a chicken-and-egg problem since drugs may never reach this point unless sponsors help. Here, open source offers a crucial advantage. Because open source collaborations depend on volunteers, they possess an inherent transparency that few biotech companies can match. As Professor Titmuss noted three decades ago, volunteers can gain nothing by shading the truth.¹⁴⁷

The case for commercial open source is similarly varied. In the case of stem cells, we have argued that open source stem cell lines would reduce input

¹⁴² Josh Lerner & Jean Tirole, *The Scope of Open Source Licensing*, 21 J.L. ECON. & ORG., 20-56 (2003), available at <http://www.people.hbs.edu/jlerner/OSLicense.pdf> (last visited July 16, 2007).

¹⁴³ *Id.*

¹⁴⁴ *But see* Opperbeck, *supra* note 13, at 171 (“Even if open source models could work in biotechnology as a practical matter, one must ask the normative question whether law and public policy *should* support such models over alternatives based on government control or privatization.”).

¹⁴⁵ Benkler, *supra* note 19, at 379.

¹⁴⁶ Solomon Nwaka & Robert G. Ridley, *Virtual Drug Discovery and Development for Neglected Diseases Through Public-Private Partnerships*, 2 NATURE REVS.: DRUG DISCOVERY 919 (2003).

¹⁴⁷ RICHARD M. TITMUSS, *THE GIFT RELATIONSHIP: FROM HUMAN BLOOD TO SOCIAL POLICY* (1972).

prices for therapeutics companies. At least in part, these savings would be passed on to consumers. This would allow more patients to purchase the drug and give society more bang for its research and development investment. On the other hand, the main advantage of open source clinical trials would be their transparency. These benefits could be extremely important if they allowed the FDA to reduce its current monitoring and paperwork costs.

Finally, all of the foregoing projects would organize science in ways that look very different from today's world of commercial secrecy and small competing academic laboratories. This would have two distinct benefits. First, volunteers would have an opportunity to pool information that might otherwise be kept secret or proprietary. In theory, companies have nothing to lose by sharing their databases with outside groups studying neglected diseases because developing world markets have little or no commercial value. However, this will only be true as long as the shared data does not leak back into the race to develop new drugs for rich nation diseases. These concerns will be much reduced if the company's employees are also open source volunteers. In this case, only group members who were already insiders would ever see the full database, although any answers they came up with would obviously have to be made public.¹⁴⁸ Second, small academic laboratories are perennially short of manpower. This forces them to replace human judgment with computation wherever possible. But human judgment is valuable. Indeed, commercial bioinformatics companies often hire large teams to sift through data by hand. A successful open source collaboration would extend this "many eyeballs" strategy to basic research.

VIII. ARE OPEN SOURCE DRUG DISCOVERY LICENSES LEGAL?

Open source software's most famous legal innovation is surely the viral license that requires improvers to offer any modified code on the same open terms as the original software. Traditionally, these licenses have been based on copyright. Perhaps surprisingly, none has ever been tested in court. Nevertheless, most commentators think that such viral restraints are valid. The situation becomes much more complicated, however, when one tries to build viral terms into a patent license. This section asks whether open source drug discovery needs viral licenses and, if so, whether such licenses are enforceable.

Are Viral Licenses Necessary? The conventional argument for viral licenses is that an initially open technology might be captured if the improvements are patented.¹⁴⁹ Alternatively, a scientist who donated technology

¹⁴⁸ Companies are much more likely to share the answers to individual search queries than entire databases. LEXIS is a familiar example of this strategy. See, e.g., Stephen M. Maurer, *Raw Knowledge: Protecting Technical Databases for Science and Industry*, (Jan. 1999) (commissioned report for U.S. National Academy of Sciences), available at <http://www2.sims.berkeley.edu/courses/is296a-3/s99/database.pdf> (last visited July 16, 2007).

¹⁴⁹ Maurer & Scotchmer, *Open Source Software*, *supra* note 105, at 303.

to the public domain might feel cheated if it later earns money for others.¹⁵⁰ These claims are far from self-evident: After all, consumers who do not want to pay for improvements can always use the original unimproved technology. Moreover, the continued existence of a public domain alternative will prevent the improver from charging more than the value of his own work – The improvement.¹⁵¹ On its face, this seems like a fair solution. Following Prof. Rai,¹⁵² one might argue that the public domain is sufficient and that some open source collaborations do not need viral licenses at all.

Nevertheless, there are at least two cases where viral terms are either desirable or essential. First, as Professors Gambardella and Hall have emphasized, many open source collaborations are unstable. Even though members collectively value a large public domain, each individual scientist can do still better by patenting his or her contributions. The result is a Prisoner's Dilemma outcome in which every scientist seeks patents and the public domain disappears.¹⁵³ This dynamic is not necessarily an argument for viral licenses since much less intrusive mechanisms like norms, peer pressure, and transparency seem to stabilize most open source software collaborations.¹⁵⁴ Nevertheless, there may be cases in which more formal legal guarantees might be needed. The second case where viral licenses could make sense is where markets tip to a single standard technology so that the existence of an open product becomes irrelevant. While tipping dynamics seem to be rare in biology, our stem cell discussion suggests that such situations will sometimes arise.

Are Viral Licenses Enforceable? Professor Opderbeck has expressed skepticism that licenses which prohibit patenting could be used to invalidate a patent or support any other “meaningful remedy.”¹⁵⁵ Other commentators have gone further by asking whether a viral license that limited licensees’ ability to

¹⁵⁰ *Id.* at 299.

¹⁵¹ *Id.* at 490-96.

¹⁵² Rai, *supra* note 18, at 148.

¹⁵³ Gambardella & Hall, *supra* note 77; *see also*, Opderbeck, *supra* note 13 at 181; Symposium on Bioinformatics and Intellectual Property, *supra* note 7 (Professor Michael Meurer comparing software and bioinformatics/genome projects and arguing that the latter pose larger incentives for volunteers to defect).

¹⁵⁴ Maurer & Scotchmer, *Open Source Software*, *supra* note 105, at 303.

¹⁵⁵ Opderbeck, *supra* note 13, at 199.

Nothing in the Patent Act would suggest that a patent could be invalidated because some of the underlying data was derived from a database in violation of the database's terms of use. Thus, it is unlikely that the HapMap license provides any meaningful remedy once a patent has been filed. Similarly, once data has been published in violation of the license terms, the horse will have left the barn. Since the HapMap project is a nonprofit venture, any damages from such a contractual breach are likely to be highly speculative.

Id.

patent improvements would constitute misuse.¹⁵⁶ The basic objection is that Congress has authorized inventors to seek improvement patents even when they are “blocked,” that is when they cannot be practiced without the underlying patent owner’s consent.¹⁵⁷ It would be strange if the parties could overrule this policy simply by signing (or clicking on) a license. Doing so, the argument runs, reduces the reward available to improvement patents and therefore innovation.¹⁵⁸ However, Professors Feldman, Boettinger, and Burk point out that this argument proves too much¹⁵⁹: if diminished rewards were the only test, then grantbacks, reach-through agreements, and patent pools would all be illegal which, plainly, they are not. Nevertheless, these arguments are not entirely convincing. Partly this is because – even in software – no case challenging a viral license has ever been litigated to judgment.¹⁶⁰ This makes arguments from existing case law highly speculative. More fundamentally, misuse cases typically involve transactions in which the right to seek patents is transferred but not extinguished. A court might justifiably ask how the analysis changes when the license blocks patent incentives entirely in favor of open source.

In the typical grantback or reach-through case, licensors and licensees agree to share improvements in advance. This can slow research and development by suppressing the incentives for the licensor and licensee to engage in (possibly wasteful) races to find and patent improvements.¹⁶¹ At the same time, the licensor and licensee still have incentives to see their shared technology improved and patented. This is obvious for the licensor, who will develop ideas for suggested improvements whenever the research and development cost is less than the expected increase in the technology’s earning power. However it is also true for the licensee, since disclosing ideas costs nothing and may persuade the licensor to invest in improvements that make the existing license more valuable. In this sense, grantbacks do not really *reduce* the licensee’s patent incentives so

¹⁵⁶ Robin C. Feldman, *The Open Source Biotechnology Movement: Is it Patent Misuse?*, 6 MINN. J.L. SCI & TECH. 117 (2004); Sara Boettinger & Dan L. Burk, *Open Source Patenting*, 1 J. INT’L BIOTECH L. 221 (2004). Both papers suggest that open source collaborations can avoid misuse by licensing data or biological materials (e.g. cell lines) instead of patents. Given the current disarray in patent misuse doctrine, this may well be true. Nevertheless, the grounds for such a distinction are far from obvious. If patent owners cannot prevent licensees from seeking and exploiting improvement patents, the same public policy objections should apply *a fortiori* to licenses based on biological materials or data. In both cases the licensee must waive her congressionally-mandated right to file and exploit improvement patents.

¹⁵⁷ Feldman, *supra* note 156; Boettinger & Burk, *supra* note 156; Opderbeck, *supra* note 13, at 200.

¹⁵⁸ Boettinger & Burk, *supra* note 156.

¹⁵⁹ See Feldman, *supra* note 156; Boettinger & Burk, *supra* note 156.

¹⁶⁰ Rai, *supra*, note 18 at 137; see also Jason B. Wacha, *Taking the Case: Is the GPL Enforceable?*, 21 SANTA CLARA COMPUTER & HIGH TECH. L.J. 451, 456 (2005); Cf. Gonzalez, *supra* note 13, at 356 (“[T]here may be a case that there cannot be a non-commercial open license for patents.”).

¹⁶¹ Incyte Pharmaceutical Co. clearly understood this dynamic when it insisted that subscribers who use its database to discover and characterize full-length genes grantback a non-exclusive license to Incyte and every other user of its data. Company council Lee Bendkey remarked that the arrangement was designed to avoid “racing” and even operated like “open source.” Hope, *Open Source Biotechnology*, *supra* note 7, at 159-60.

much as *transfer* them to the licensor. Since no incentives are destroyed, we expect all profitable ideas to be implemented sooner or later.

This same logic would apply to the de facto patent pool established by Cambia's protected commons scheme. As long as the pool's size was limited, existing members could be sure of recapturing all profits from any improvement. Knowing this, we would expect them to disclose profitable ideas and share development costs as necessary. As in the commercial case, Cambia's grantback clause would only redistribute patent incentives, not destroy them. The answer would be very different, though, if Cambia designed its commons so that membership really was "accessible to all."¹⁶² In this case, the defining feature of patents – the power to exclude – would no longer apply. Instead of transferring patent incentives, Cambia's grantback clause would make them disappear entirely.

This result might be acceptable if we knew that new ideas would continue to be developed anyway under open source incentives. However, there is no reason to expect this: In general, we expect patent incentives to support at least some R&D projects that open source does not.¹⁶³ For this reason, a Cambia-style viral contract potentially poses much greater dangers than the grantback clauses found in conventional patent pool cases like *Transparent-Wrap*. We should not lightly overrule Congress's policy judgment that patent incentives are an important research and development incentive. At the same time, it makes sense to develop ideas using open source methods as much as possible, since this means fewer patent monopolies. It therefore seems reasonable to let parties write viral licenses for limited periods of time so that patent incentives can be brought back into play if open source methods fail. The HapMap Collaboration's decision to waive license restrictions that had initially prevented users from seeking patents is very much in this spirit. Future open source drug discovery collaborations would be well-advised to design viral terms that fade away over time.

IX. CONCLUSION

In the late 1990s, the idea of open source drug discovery seemed so inevitable that there was no particular need to ask how it would work. "Wait and see" seemed like a sufficient response. Today, commentators have fallen into the opposite fallacy of dismissing the concept in a few words. We have avoided both extremes, arguing that open source drug discovery is neither inevitable nor impossible, but merely difficult. In the process, we have identified five scenarios where open source ought to work and would be a distinct improvement compared to conventional patent incentives. Nobody would claim that this list is exhaustive, but it does show that open source drug discovery is worth thinking about. Additional examples are limited mostly by our imagination.

¹⁶² This could be done, for example, by eliminating the "subscription fee," in-kind contributions, and "member approval process." See *supra* text accompanying notes 81-89.

¹⁶³ Maurer & Scotchmer, *Open Source Software*, *supra* note 105, at 288-90.

Imagination, of course, is what commentators are supposed to supply. Over the past decade legal scholars, economists, and other social scientists have learned a great deal about the incentives that drive ordinary open source software. For this reason, the idea of open source drug discovery comes as a sort of final exam – similar to software in many respects but also sufficiently different for naïve analogies to be misleading. Success will require a much deeper and more detailed understanding than anyone would have imagined ten years ago.

The stakes are high. Open source is often a plausible strategy for reducing drug development costs and making new medicines affordable; it offers increased transparency for funding agencies trying to decide which early stage drug candidates to invest in; it may allow regulators to reduce the reporting requirements that help make late-stage drug discovery expensive; and its ability to mobilize volunteers offers a key advantage to cash-strapped neglected disease programs. Perhaps more importantly, open source is the first fundamentally new innovation mechanism since patents and copyright appeared four centuries ago. A successful drug discovery collaboration would show that open source methods are not limited to software and could even open the door to applications beyond biology.¹⁶⁴ For all these reasons, open source deserves a second act.

¹⁶⁴ The computerized map (“Geographic Information Systems”) industry provides an especially promising venue for open source methods. Map users routinely find that streets have been mislabeled or do not exist. Open source methods are an effective way to find and eliminate these bugs. NATIONAL RESEARCH COUNCIL COMMITTEE ON LICENSING GEOGRAPHIC DATA AND SERVICES, LICENSING GEOGRAPHIC DATA AND SERVICES (National Academies Press 2004).

Table 1: Common Open Source Incentives¹⁶⁵

| | |
|--|---|
| <i>Own Use (Personal)</i> | Writing software for one's own use and enjoyment. Examples include hobbyists, developers (LINUX). |
| <i>Own Use (Corporate)</i> | Producing an open source product used by one's employer. Examples include corporate webmasters' creation of the Apache web server. |
| <i>Education</i> | Producing an open source product in order to learn from experience and peer review. |
| <i>Signaling</i> | Producing an open source product in order to demonstrate competence to others. |
| <i>Shared Research & Development Costs</i> | Firms that do not compete with one another frequently use open source to share research and development costs. For example, companies that share web development costs (Apache) lose little, if any, competitive advantage by doing so. |
| <i>Related Goods & Services</i> | Producing an open source product which is needed to sell a separate, proprietary good or service. Examples include hardware (IBM) and customer training and programming services (Red Hat). |
| <i>Social Psychology</i> | Working for nonmaterial rewards including reputation, altruism, collective solidarity, etc. |

¹⁶⁵ Maurer & Scotchmer, *Open Source Software*, *supra* note 105, at 287–300.

Table 2: The Drug Discovery Pipeline¹⁶⁶

| Task | Description |
|----------------------------------|--|
| <i>Basic Research</i> | Undirected, curiosity-driven research into the mechanisms that cause disease |
| <i>Finding Targets</i> | Exploiting research to find a gene location, metabolic pathway, or other point where drugs can intervene to disrupt disease |
| <i>Validating Targets</i> | Using multiple, additional lines of evidence to see whether they support or discredit the hypothesis that a given target can be used to disrupt disease |
| <i>Finding Lead Compounds</i> | Testing chemical compounds to see whether they bind to an existing target and otherwise possess the properties needed to make an effective drug. Lead target searches can be conducted using “wet chemistry” methods or computer simulations |
| <i>Optimizing Lead Compounds</i> | Systematically modifying and testing lead compounds to increase their effectiveness as drugs. |
| <i>Process Development</i> | Developing procedures for making candidate drug in large, affordable quantities |
| <i>Pre-Clinical Testing</i> | <i>In vitro</i> and animal testing to determine candidate drug’s safety and efficacy |
| <i>Phase I Tests</i> | Testing the candidate drug in 20 – 80 healthy volunteers to find safe doses and identify side effects |
| <i>Phase II Tests</i> | Testing the candidate drug on 100- |

¹⁶⁶ Stephen M. Maurer, *The Right Tool(s): Designing Cost-Effective Strategies for Neglected Disease Research*, (2005) (commissioned report for World Health Organization), available at http://gspp.berkeley.edu/ihs/MAURER_DrugRsch.pdf (last visited July 16, 2007).

| | |
|------------------------|---|
| | 300 patients to obtain short term safety information and preliminary efficacy data |
| <i>Phase III Tests</i> | Testing on several hundred to several thousand patients to search for rare side effects, document efficacy, and optimize delivery methods and doses |
| <i>Approval</i> | Paperwork and hearings to obtain FDA approval of the candidate drug for specified uses |
| <i>Phase IV Tests</i> | Testing on patients to confirm pre-approval test conclusions; demonstrate drug's efficacy in applications not previously approved by FDA |

