# SYMPOSIUM ON BIOINFORMATICS AND INTELLECTUAL PROPERTY LAW

## APRIL 27, 2001—BOSTON, MASSACHUSETTS

## OPEN SOURCE GENOMICS

*PROFESSOR MAUREEN O'ROURKE*:

Our last panel of the day features Dan Burk, who is a professor at the University of Minnesota Law School, and he is also the Vance K. Opperman research scholar at University of Minnesota. Professor Burk holds appointments at both the law school and the center for bioethics, and is also the Associate Director for the new joint degree program in law, health and the life sciences. Before his arrival at the University of Minnesota, Professor Burk taught at Seton Hall in New Jersey. Without further ado, then, Professor Dan Burk.

*PROFESSOR DAN BURK*:

Thank you. Well, I am glad to be here, and I have enjoyed the conference so far. I want to change gears a little bit with my topic but try to get us to think about some of the themes that have run throughout the presentations today. I would like to go back to thinking about how innovation works, how research gets done, and what the role of intellectual property rights might be in generating the kind of bioinformatic information and data that has been the focus of discussion today. One point runs through the previous presentations that we have had that but has not been addressed. That is that as biotechnology begins to move into the bioinformatics area and as it becomes a tool and an issue, we are seeing the merger of the two areas that tend to get talked about as being "high technology" innovation: the biotechnology area and the informatics or digital revolution.

Those who have watched each of these areas know that they have moved at very different paces for quite some time. The pace of innovation and change in biotechnology has been relatively slow, in part because of the long development times and the product cycle in biotechnology. In the area of cyberspace or the digital revolution, things have been moving very rapidly. In fact, some of us at dinner last night were talking about the milestone events

that took place in the good old days of cyberspace law. We realized that these events we were talking about took place only about two years ago.

So the pace of change in informatics has been very rapid, and as biotechnology and informatics become married, we are going to see more and more what we have heard about in the last set of presentations: we will see the acceleration of the pace of change in biotechnology, and the question I want to ask is, can we learn some things in the biotechnology and bioinformatics area from the scholarship and discussion that has been taking place in the rapidly-evolving area of cyberlaw and digital technology? I am going to suggest that the answer is yes. There are some interesting parallels and contrasts, and some of the scholarship that has been developed in one area can be carried over into the other. I titled this presentation "Bioinformatics Lessons from the Open Source Movement." That leads me to talk for a minute about what the open source movement is, for those who may not have focused on that phenomenon in the area of software and digital rights.

The so-called open source movement is sometimes called the free software movement. Those terms are sometimes used interchangeably, sometimes they refer to slightly different communities, but they are generally used to refer to communities of programmers who are committed to certain principles in terms of the way they work and the software they produce and the way that software is used. The phrase that Richard Stallman uses is, "Free software is not like free beer." The term "free software" does not refer to software that you get for free or at no cost. Rather, it refers to the ability to manipulate and to change programs that are written, to use them in certain ways, and to make certain that people further downstream can continue to use them that way as well. One of the tenets of this movement is that the source code that will allow you to understand and change and manipulate the software should be freely available, which is typically not the case with most commercial packages, which are distributed as object code.

Typically the software that is produced by open-source programmers or free software programmers is accompanied by some sort of licensing agreement that is designed to keep this freely available to other users downstream. The most famous of these which you may have heard talked about in the media is the so-called GNU Public License. GNU is an acronym for "Gnu's Not UNIX," which is the type of operating system that is produced using this open source methodology. There is a little recursive joke to it. It is accompanied by a public license, and there are other versions of these licenses, but they tend to have in common certain features. Some people refer to this as a copyleft, as opposed to copyright, license. The features of copyleft are that you are allowed to take the program and its source code and modify it under the condition that you allow other people further downstream to continue to modify it and that you make your source code available as well. That is why they refer to it somewhat jokingly as copyleft. Rather than trying to assert proprietary rights to keep that code in one particular form and keep it away from other people who might modify it, copyleft uses proprietary rights for the purpose of keeping the program open and available and free to people who

might modify it.

There is a very striking sort of parallel when you think about the tenets of this open source movement that is committed to the free flow of information, and the tenets of the scientific research community. Many of open source programmers work as volunteers providing their work as a public service. We have seen some of the same tenets of this movement talked about in the literature with respect to the Human Genome Project and similar types of scientific research projects. Both share this sort of distributed production model in which you have many small units of production that are not, as far as is immediately apparent, centrally coordinated. Each of them has a normative structure that is committed to this idea of the free flow of information and the commons and sharing the output communally. Rebecca Eisenberg has written about this in the scientific research area.[1] Arti Rai has also written about this a little bit,[2] and they show that some of this behavior in the genomics area is captured in the work of Robert Merton.[3] Merton has been criticized for not getting everything right, but I think anyone who has worked in a biotechnology laboratory still feels some of those norms that Merton talked about: that you are supposed to share data freely not to work so much for commercial rewards as for reputational rewards, etc.

One sees many similar statements about the norms of this open source programming community, the so-called "hacker ethic," committed to the free flow of information and sharing their output, and working, again, for reputational rewards rather than for monetary rewards. Both communities seem to be very concerned about the capture or commercial conversion of their output. Articles in *Science* and *Nature* and so on indicate the concern of academic researchers and the scientific community that Celera or other types of commercial firms might capture this information that they are generating. You heard references earlier today about attempts to try and keep that from happening by creating prior art so that patents cannot be filed. You see the same sort of concern in the open source community's norm against commercialization, against proprietary capture of the software that is produced, particularly concern that Microsoft is somehow going to capture this code and turn it into yet another part of Bill Gates's dominion over the free world.

Thus, there are some very similar sorts of norms and concerns in each area. I want to focus on the public domain capture problem for just a minute, because this has been talked about a good deal in the digital rights literature. Yochai Benkler[4] and now Larry Lessig[5] have talked about different levels at

---

[1] *See* Rebecca Eisenberg, *Proprietary Rights and the Norms of Science in Biotechnology Research*, 97 YALE L.J. 177 (1987).

[2] *See* Arti K. Rai, *Regulating Scientific Research: Intellectual Property Rights and the Norms of Science*, 94 NW. L. REV. 77 (1999).

[3] *See* Robert K. Merton, *The Normative Structure of Science*, *in* THE SOCIOLOGY OF SCIENCE: THEORETICAL EMPIRICAL INVESTIGATIONS (Norman W. Storer ed., University of Chicago Press 1985).

[4] Yochai Benkler, *From Consumers to Users: Shifting the Deeper Structures of*

which you might have the kind of openness that the open source movement is committed to, or conversely different levels at which technology that is available to the public might be captured and made proprietary. They talk about the physical layer, the logical layer, and then the content layer. If you think about the telephone network or the Internet, the way they function now, you can quickly see these different layers. Information flows over physical wires that are owned by somebody, frequently by AT&T or your cable carrier or someone like that. There is a logical layer that routes and directs the information, and in the case of the Internet, that is open and available to everyone; the Internet protocol is not proprietary. Then, riding on top of that, you have content that is sent over those wires using those protocols, and the content may be copyrighted and proprietary or it may be public domain and nonproprietary. Thus, you could have any of those layers either open or proprietary, and to the extent that any of those layers is captured or made proprietary, that particular layer, at least, would be unavailable to the public.

There is a great deal of concern at each of those levels about the possibility of capture of the Internet by Microsoft or AT&T or whomever. The same is true if we think about the bioinformatics realm we have been talking about today. There is going to be some hardware level, though we have not really thought about that much or talked about it today. There is going to be some logical level to that bioinformatics information. We have heard some discussion here of nomenclatures, of normalization, of standardization, of indexing, etc. Finally, there is the data, the content itself. There could be proprietary rights at any of those levels. With respect to each of these levels, we are going to see what economics talks about as network effects. The telephone system is the paradigm example of this: once you have the telephone system in place, it does not cost much to add another user, but when you do, the value of the network most certainly goes up to the people who are already on the telephone system. If you only have one or two people on the telephone network, it is not very useful or valuable, but if you have hundreds and thousands of people joining it, each new user that joins increases the value of it. The same is true of other types of real and virtual networks, and the Internet and certainly bioinformatics would fall into that category.

We have heard discussion here today about standardization and the problem that Dennis Karjala mentioned of trying to put different protocols or types of databases together. If those are standardized, interaction will be much easier, and the information will be much more valuable. The problem that we know from the economic literature is that if you believe there is such a thing as a network effect, if the standard that becomes settled upon is proprietary, you get a kind of lock-in that may be undesirable. We can talk about that during the question and answer period. There is this problem with tipping, which is what

---

*Regulation*, 52 FED. COMM. L.J. 561 (2000).

[5] LAWRENCE LESSIG, THE FUTURE OF IDEAS: THE FATE OF THE COMMONS IN A CONNECTED WORLD (Random House 2001).

the *Microsoft* case[6] was all about: Can people with intellectual property rights influence or push the standardization process in a direction that benefits them, and if they do that, is that desirable? Those are things we might want to be worried about in the bioinformatics area. Certainly there has been concern about it in the digital technology area. If there are certain standards or conventions that are settled upon for bioinformatics, either in terms of the content or the indexing or the logical layer, would we want those to be proprietary standards? Would we be concerned if people pushed things in a direction so that their proprietary standard became the one that was generally adopted?

There has been an evolution of the literature in the open-source and digital rights area, talking about this movement in terms of the theory of the firm. I want to spend a minute or two to lay that out for you because it illuminates this question and concern that we might have in bioinformatics about proprietary standards and the sorts of norms and concerns that each of these two communities share. The question here is simply, when you have these communities like the open source community or the human genome sequencing project that seem to be very atomistic and distributed and lacking in central coordination, how does the work get done and why does it get done? Would it be better done in a traditional commercial sort of firm? Certainly Celera seems to think that that is the case. In the software area, Microsoft or even Red Hat would seem to think that is the case, and yet a lot of work does get done. How can that be the case?

Coase talked about the firm.[7] This is one of his major contributions to economics and to our understanding of the law. He talked about the firm as an area of economic activity that is organized in a certain way, hierarchically rather than as an open market with market transactions, directed by some central authority, an entrepreneur, where these interactions are not negotiated but they are much like command and control. Other people took that a little further and said, well, the size of the firm and the scope of the firm will be determined by, first of all, the transaction costs. It is costly to use markets, and if it is sufficiently costly, it might be better to have this area of activity that is organized hierarchically, but it is also going to be bounded by the ability to coordinate efforts within that firm and also by agency costs. You have many different people acting within the firm, and some of them might have interests that are not well aligned with the interests of the firm.

You can also think about the firm as a nexus of contracts. We say it is hierarchically organized, but part of that is by contracts that could be employment contracts or other kinds of production contracts, depending on how vertically or even horizontally integrated that firm might be. These contracts define the kinds of things that happen within that economic space,

---

[6] United States v. Microsoft Corp., 253 F.3d 34 (D.C. Cir. 2001).

[7] Ronald H. Coase, *The Nature of the Firm, in* THE FIRM, THE MARKET, AND THE LAW 33 (University of Chicago Press 1990).

and as I said, agency costs are going to be a problem. You may have employees who have interests that differ from that of the entrepreneur who is directing things in the firm. How do you get those employees to fall into line? Well, the employment contract is one thing. You do not pay them or do not give them certain rewards if they step out of line. There may also be other kinds of bonding or coordinating mechanisms to cut down opportunism. In the software area, we have seen the use of stock options. If the firm does well, the employees are richer, and if the market tanks like it has, the employees are poorer but we hope they will pull together to make those stock options worth more than they would have otherwise.

How about intellectual property rights? Can those be used in these ways to make the firm operate better? Well, you heard Arti Rai talk about some of Ed Kitch's work in the so-called prospect theory of intellectual property rights,[8] and one way to think about that is that it is a type of coordinating mechanism, sometimes between firms, but even within a firm. We give an intellectual property right to one particular entrepreneur who then coordinates and directs the development of that particular resource, and that was how Kitch thought we should view intellectual property rights. There are also intellectual property rights that are bonding mechanisms. The classic example here might be something like trade secrecy. In order to keep my employees headed down the same path I want to go down, I might put some legal restrictions on how they use certain information and what information they can take with them when they leave the firm. That raises the question, then, if intellectual property can be used to coordinate development of certain resources, do we see that happening in these communities that do not look very much like traditional firms? In the open source community writing software, and with the Human Genome Project trying to put together a picture of what the human genome looks like, we see intellectual property being used to try and help coordinate those efforts.

Now, a lot of the coordination in that community is clearly normative. There is a certain expectation that you will write code if you are part of this community, you will share that code with other people, and you will not work on projects that someone else has staked out as their territory. There is still concern, however, that someone will defect, will try and commercialize one of these software packages and will try and take it away from the commons. This GNU public license that I mentioned, this idea of copyleft, is used to prevent that from happening and also to coordinate the development of these software packages. The most famous example of such a package that is accompanied by one of these licenses is Linux. The copyleft license that accompanies it is supposed to keep people from defecting and turning a modified version of the software into a proprietary package.

---

[8] *See* Edmund Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265 (1977).

Now curiously enough, over in the human genomics side, there has been discussion of using intellectual property rights, but that has been either in a very negative sense or it is been largely rejected. Some of you may recall that Rebecca Eisenberg talked earlier today about the NIH patents that were filed on all these expressed sequence tags when Craig Venter was still part of that effort. Reid Adler, who was the technology transfer guy who did that filing, said the reason they filed these patents was to prevent people from staking out claims or capturing information. They wanted the government to do it and essentially used the patents to keep the information open and free. The NIH EST patents caused an enormous outcry among the scientific research community, and eventually the NIH stopped trying to prosecute those patents. There has been a fair amount of pressure not to go back there, and the patents have been filed instead by the commercial firms, with some degree of disdain from the rest of the research community.

Whereas the open source software coding community has used the copyright and the license of that copyright to prevent capture, patent control was an option that was rejected in the genomics area. That means that people who do want to move into this area and commercialize this information are operating in very different licensing environments. A company like Red Hat, which wants to commercialize and have a more traditional business model using the Linux software package produced through this open source effort, is unable to do certain things with Linux software. In fact, in the disclosure they filed with the Securities and Exchange Commission, they talked about the fact that they do not really have control over the product they are selling since it is accompanied by this license which requires them to make public the source code of any changes that they made to Linux. Red Hat's product is subject to this copyleft or GPL, and they are somewhat constrained in what they do in their business model. With a company like Celera, however, you see something very different. There is been an effort, as was mentioned earlier today, to try to prevent commercial patenting of gene sequences by throwing information into the public domain as quickly as possible. Of course, Celera and Incyte and other pharmaceutical firms that specialize in providing these databases sucked that public information right up and made it part of the product they already have. As a result, they are relatively unconstrained by intellectual property rights that people might have in the way they want to use the information.

This difference in approach to using intellectual property rights to try and coordinate developments has resulted in two very, very different environments for commercialization. We may think that is a good thing or a bad thing depending how you view commercialization of information, but it clearly has led to different outcomes. A bit of different take on this – and I apologize to the commentators because this is a very recent development in thinking on this – is present in the literature that has grown up in the open source area on so-

called actor-network theory.[9] I mentioned a few minutes ago the work that Robert Merton had done on the normative structure of the scientific community. Things have advanced since then, and his work has been criticized, as I said, and people have advanced other sorts of theories as to how science really operates.

One view principally put forward by Bruno Latour postulates that really anything in society, but science in particular, operates as a series of relational networks,[10] and this parallels, in some sense, Coase's theory of the firm that I mentioned a moment ago. These networks define areas of effort and contain both human and technological components that people call actants. They are not really actors because some of them are bits of technology rather than humans. But a scientist, for example, is only a scientist because she has a laboratory and centrifuges and test tubes and Eppendorf tubes, etc., that really define that particular role as a scientist. Part of this theory of these relational networks discusses something called blackboxing, which is sort of a social shorthand. We talk about someone as a scientist, but we do not try to describe or look behind what that means. We are sort of vaguely aware that there is this network of training and universities and government funding and centrifuges and workbenches and so on back there, but we just talk about it as a "scientist," and that is sort of a black box into which we do not look.

This is true of other types of entities as well. We talk about Microsoft as a firm, and there are really lots of employees and machinery and all sorts of inputs there within that particular black box. We have talked today about the open source movement, which is a sort of black box, or the Human Genome Project, which is a black box full of all sorts of actants and relationships. When you blackbox something like that, Latour says you suddenly have a control point for whatever is in that box, depending on what the dimensions of that box are and how you define that particular label that you attach to it. One of the ways we might think about the idea of capturing this information that would otherwise be out in the public domain is as blackboxing it. A software package or genomic information could be incorporated into one of these other black boxes or one of these other relational areas that can then be marketed and sold. This blackboxing process involves control of some of these relationships that are there.

Intellectual property has been discussed in the open source area, and it is an important part of that blackboxing phenomenon. We talk about someone being subject to a patent or a work that is copyrighted, and that is part of what is going on in the genomics area just as it goes on in the open source area. One can take information, put it into that black box or put a label on it, and then you have a type of control over it because of the copyright or the patent or the

---

[9] *See* Ilkka Tuomi, *Internet, Innovation, and Open Source: Actors in the Network*, FIRST MONDAY, *at* http://www.firstmonday.org/issues/issue6_1/tuomi/index.html (Jan. 8, 2001).

[10] *See* BRUNO LATOUR, SCIENCE IN ACTION: HOW TO FOLLOW SCIENTISTS AND ENGINEERS THROUGH SOCIETY (Harvard University Press 1988).

mechanism of blackboxing that you would not have had previously. This is an area where it seems there is some fruitful cross-fertilization potential between software and the open-source movement and bioinformatics.

My presentation probably leaves more questions hanging than I have answered. This is a work in progress that I wanted to share with you, but there are a couple of take-home questions to which we might want to give some thought. Why has there been or is it a good thing for there to have been this divergence in approach between these two communities with regard to intellectual property rights? Would it make sense, given some of the concerns that have been raised here today, for there to have been something like copyleft or GNU Public Licensing for information that is put out into the public domain or the bioinformatics area? There are probably a number of other questions that we could ask about that, but I will leave those for the question and answer period and turn some time over to the commentators.

Thank you. (*applause*)

*PROFESSOR O'ROURKE*:

We have two commentators today. First is Josh Lerner, who is a professor of business administration at Harvard Business School and also a Faculty Research Fellow in the National Bureau of Economic Research, Corporate Finance, and Productivity Program. After Josh, we will conclude, fittingly, with our conference organizer, Professor Michael Meurer of Boston University School of Law. We owe him a debt of gratitude for getting the conference together, and we at BU are delighted that he chose to join us a couple of years ago from the University of Buffalo. Without further ado I will turn it over to Josh.

*PROFESSOR JOSH LERNER*:

Thank you. This presentation is a little bit challenging to discuss. Since I had only a short summary to review in advance, I decided to discuss the paper that I *thought* Dan was going to write. (*laughter*) I think at least I anticipated a big chunk of the talk, and while I may have missed a few things, but I hope that we will have an interesting discussion.

There are two parts of the paper. The first part presents the argument that there are many parallels between open source and biotechnology research or, more generally, academic research. The argument Professor Burk made here was strong and quite persuasive. The area I was less comfortable with was the argument that free-riding problems, which apparently the author believes are avoided through the GPL or related open-source licensing regimes, could be addressed by a similar licensing regime in biotechnology.

We can enumerate many different features of open-source and academic research where there are parallels. First of all, there is the nature of the innovation process itself: both communities emphasize disclosure, getting the stuff out there early. Early on, there are big rewards not to hold things back,

but instead to build up a reputation by moving early and generating contributions. People are aware of these contributions and very keenly sensitive as to who is making which contributions.

Furthermore, there is similarity in the distribution of the output. In academia, there are a few "rock stars" that generate a disproportionate share of discoveries. So too, there is an enormously skewed distribution in open source, where there is a relatively small number of programmers who do the great deal of contributions.

We also can think about the similarities in the nature of the governance system. In particular, as Dan suggested, open-source communities, like academia, are not command-and-control type environments. There is leadership in these projects, but it is a different style of leadership. In particular, one of the big concerns in these open-source communities is that the project's founder cannot prevent people from taking the project and going off some other way that he intends (which is typically called "forking" in this literature). It is just as a grand old man in this field of an academic field cannot say, "Do not do research in Area A, but instead look at Area B." In these fields, leadership is not a matter of formal control, but one of persuasion or informal authority.

Similarly, there is the process through which non-commercial contributions and career concerns interact with each other in both arenas. While this is an issue that open-source people do not always like to acknowledge, many people in open-source, just as many academics, have managed to translate their reputation and their stature in the open-source community into commercial gains. A large number of open-source pioneers have raised venture funding for companies they have set up or gone to work for large companies, just as professors find their way into start-ups and consulting opportunities.

Finally, there are some of the same distortions in academia as in open-source. In particular, one of the arguments that we made in our research about open-source[11] is that you can have some of the same problems with fads that you see in academia. If you are contributing to something because you want to be in an area where you can develop a reputation, you are going to choose a field where you think there are going to be a lot of contributions later on. Just as we see in academia – where there are certain topics on which it seems every graduate student wants to write a paper because they know there are a lot of people in the profession that are interested in the subject – so too it seems there is some of the same reinforcing or faddish behavior on the open-source side. As a result, this behavior can lead to too great similarities in the types of projects undertaken.

For all these reasons, I am comfortable with the author's parallels between open-source and academia. Where I feel less comfortable is with the implications that are drawn from the comparison. Particularly, there is a quote

[11] Josh Lerner & Jean Tirole, *Some Simple Economics of Open Source*, J. INDUS. ECON. (forthcoming 2002).

from the paper that says, "The genome sequencing community rejected legal mechanisms chosen by the open-source community to enforce its behavioral norms. This opened the way for commercial firms to free-ride."

The questions I really have are twofold. First of all, when we look at what is happening in open-source today, to what extent are the GPL and some of the other related licenses really effective in avoiding the kinds of free-riding problems that the author is worried about? Second, when we think about what biotechnology is and the nature of biotechnology versus the nature of software, does this logic really carry over? I will talk about each issue briefly.

In particular, one sees a whole surge of efforts today to commercialize open-source projects.[12] First, traditional hardware or service companies like IBM are bundling their own software and their own services with the open-source code. There are cases in which founders of open source projects, like Eric Allman at Sendmail, have basically set up a commercial company that operates alongside the open-source project. We even have efforts by companies setting up exchanges in which they are getting programmers to contribute to projects over which they retain proprietary control or special rights, such as in Collab.Net.

Certainly, viewed from the open-source purist's point of view, a lot of these developments are really problematic. One of the dominant features of the GPL is its viral nature. Basically, if you bundle together software that is open-source with proprietary software, the proprietary software is supposed to be then open too, and you are supposed to make the code readily available. If you look at some of what some firms are doing in terms of taking Linux and building products around it, it seems they are very much crossing that line. Furthermore, there are many examples of commercial entities moving away from the GPL to other licenses. Hewlett-Packard, for instance, is licensing its software using licenses that give it a great deal of power to reach in and grab improvements to its software and privatize it once again.

One interesting question is why is it that many of the open source advocates are not screaming and litigating over this? I think the answer is two-sided. One perspective is that there are real questions as to whether GPL is really valid, because its clauses have never actually been tested in litigation. People who are advocates of copyleft do not necessarily want to see the validity of these licenses being overturned. It is sort of akin to our Taiwan policy before George Bush bumbled. (*laughter*) There is something to be said for strategic ambiguity. On the other side is the whole problem of "capture": many of the people who are the pioneers of open-source development are now raising venture capital or working in special consultative relationships with corporations and so forth.

A related concern with Burk's implications is the very different significant differences between software and biotechnology. When you look at Eric

---

[12] *See* Josh Lerner & Jean Tirole, *The Open Source Movement: Key Research Questions*, EUROPEAN ECONOMIC REVIEW, at 819-826 (2001).

Raymond's writings,[13] one of the key things that he believes makes open source work is this immediate applicability: that users, in his phrase, can "scratch their own itch." As a user, you can take the open source code and use it to solve problems right away and have immediate applicability. It is not clear, when we look at a realm like biotechnology with extremely long lags and large costs to commercialize products, that this essential element is present.

Just to wrap up, the discussion here is really provocative and suggests a lot of interesting issues. I certainly concur that there are all these parallels between academia and open-source. At the same time, however, the broader extensions, especially the claim that biotechnology should be adopting some GPL-like licensing, are ones I am less comfortable with.

Thank you. (*applause*)

*PROFESSOR MICHAEL MEURER*:

Well, we've made it to the end of the day. I am very appreciative of the performance that everyone has given today. Thank you.

I am going to make a few comments in the spirit of those made by Josh. I think the topic of Dan's paper is very interesting. It covers a lot of ground, and I do not have time to talk about more than a single issue. Josh's began discussing the similarity between academic research and open source, and then analyzed GPL. I want to go back to the first issue Josh raised and argue that, unlike Josh, I see some fairly significant differences between open source software and genomics. I want to talk about the cooperative ethic in the open source movement and compare it to the cooperative ethic in bioinformatics and argue that I do not see the ethic being as sustainable in the genomics world as it is in the open source world.

The first point I want to make is that history matters. I do not know if we should look at Dan's talk as a lament or a prescription. It might have been possible if we went back to 1990 and started from that point, but given our history over the past 10 years, it would be a tough row to hoe. It is tough to reverse things when we have a lot of people that are committed to an exclusive and proprietary approach. Secondly, I want to talk about incentives that bring people to cooperate in a situation in which they might free ride. My comments are based on a paper that Josh Lerner and Jean Tirole have written about the open source movement, so I am borrowing a lot from Josh's work, and he therefore cannot tell me I have it wrong.

Josh talks about how the positive incentives to cooperate include, first of all, a career concern. You can build a reputation for being a star and then be an attractive employee. Secondly, there is a norm of sharing. Working against that norm in the open-source world, and more strongly in the biotechnology

[13] *E.g.*, ERIC RAYMOND, THE CATHEDRAL AND THE BAZAAR (O'Reilley & Associates 1999).

world, though, is the possibility of cashing in on property rights, and for an academic, getting tenure. One factor that is considered in tenure decisions is how many patents have you filed. That is also a factor determining what kind of lab support you get, how many grants you get. Thus, the incentives might be a little bit stronger in the academic life sciences to grab the prize rather than continue to contribute to the public good.

The way I think about it is in terms of game-theoretic models of cooperation. Economists long ago puzzled over why OPEC was so successful. A lot of work was done in the 1980s to explain how we can achieve cooperation in situations in which people would be naturally inclined to cheat – cheat, for example, on the rules of OPEC to undercut prices that prevail in the cartel. The answer is that you could cheat and profit handsomely for a short period of time, but you will lose cooperation in the future. You might see punishment directed specifically at you, so that you balance off the short-term gains from cheating against the long-run loss. That is what sustains OPEC. It is that kind of trade-off that we want to think about when we look at situations like open-source software or academic science and consider whether we can sustain in a non-cooperative environment free of strong property rights, free of integration into a single firm, a rich level of cooperation.

I want to talk about two contrasts. First, I think there is a bigger defection incentive – and Josh was touching on this in his closing comments – in the bioinformatics situation. Secondly, a technical concern to economists when they look at repeated games is whether there is a capital T, whether the "world" is going to end or whether it will continue on for a long period of time. The OPEC model of cooperation supposes that Saudi Arabia, Iran, Iraq, all the countries that are members, think they are going to be cooperating for an indefinite period of time. They do not see the end of the game coming at any particular point. On the other hand, if you think the end of the game is coming at any particular point, then there is a kind of backwards induction that leads to unraveling and undermines the possibility of cooperation. Nobody wants to cooperate in the last period, and they start moving back in time and nobody cooperates at the beginning either.

The reason I think a lot of bioinformatic projects are different is that you are moving along to a target, to a drug, and the game is over at that point. With software, it is continually mutating. New features are added; it is moved to new platforms; there are more efficient routines, etc. I do not see permanent, perpetual use of software products, but I also do not see, as a member of the open-source movement, a particular date when the development of Linux is coming to an end either.

I want to push further with that point, and my crude understanding of open-source software and genomics might be a liability, but at any rate . . . It seems as if there is no pot of gold for anyone to grab for at any particular point in time in the software creation process. There is no point in time where you would say this is a great time to defect, because the gains are going to be very large today. I know I will lose something in the future, but I just cannot resist the temptation. I see that as being more of a problem, though, in the

bioinformatics world or the genomics world, where there will be these pots of gold that people see are ripe to be seized. That is more likely to create an irresistible temptation to defect. The basic difference in the genomics process, the bioinformatics process, is that step by step, we are adding depth to research tools, and we are producing more information that is going to be helpful in finding an end product, but it is that end product where we cash in on the market. That is what is going to create the risk of bailing out at a particular point in time.

That is only a slice of the issues that Dan is talking about in a very interesting paper, but that is where I guess I get off the bus at the first stop with regard to the similarity between open-source and academic science.

Thank you very much. (*applause*)

*PROFESSOR BURK*:

I found most of these comments very helpful. Let me start by just replying briefly to Mike's comments. They surprised me a little bit. I think you are right that history does matter, and as people like Rebecca Eisenberg have talked about, we have this very long history in the academic research community of supposedly adhering to these norms of communality and sharing and so on. It is not immediately clear to me, however, that you are going to know when the last round is in that situation either. These people are, for example, in the genome project, engaged in incrementally adding to this large database, and when we are done with the genomics, then we move on to the proteomics. Thus, it seems to me that there might be much larger temptation to defect in the open-source community where the product times, as we said, are very short. We know what Linux is worth right now, and as Josh pointed out, you can use it immediately. We are not going to have to wait for drug development, and there seems to be a very big payoff available immediately. Where the community is actually one that is relatively young, it does not have much of a long or entrenched history with the kinds of mechanisms that we see in the scientific community like journals and peer review and so on that have grown up. I think your point is well taken, but I might actually come out the other direction, taking the signals to indicate there is more change of defection in open source.

*PROFESSOR MEURER*:

I would like to respond. You characterize the cooperative problem differently than I did, and you are looking at lots of people who are just in the business of doing life science research, and they will stick with it for a long period of time. Instead I was looking at particular projects with definite end points. When you look at open-source, when you look at Sendmail, you have a definite project and community people are working on that – Linux is probably the best example – so I am not sure which is the better way to look at it. I appreciate your point. One difficulty with your point, though, is that it

becomes harder to identify the community when you have lots of projects with different players. The identity of the players can be different on different projects, and that creates a new problem for sustaining cooperation when people are coming and going. If the project is specific and the players are the same, then cooperation is easier to sustain.

*PROFESSOR BURK*:

I think that is right, and one of the thoughts that occurred to me during your presentation is that, a little bit like antitrust analysis, it depends on how we define the market. It is going to depend on how we define the project and what we think the likely defection points might be. I am not entirely certain what those are, but I am not sure that I would intuitively think that they are what you immediately represented. I found Josh's comments to be equally useful. In part, I had thought about this problem of immediate applicability. At least some of the bioinformatics information we are talking about, some of the data, is going to be immediately applicable. As Incyte and Celera have proven, I can take this information and immediately license it out to people who are doing the end product research. Now, that is not as good as getting the end product itself, but there is some applicability there.

Again, I am not sure; it depends on which product we are talking about. Is the information itself the product, or is it an input into a downstream product that is going to happen later on? Let me just mention one problem that I thought Josh was going to raise but did not, which bears some more thinking. Josh talked about enforcement. In the open-source area there is at least the threat of enforcement by the Free Software Foundation or by Evan Moglen or other folks who are wandering around supposedly anxious to sue people if they violate the GPL, although I agree about the strategic ambiguity point. It is not clear who would perform that function if the scientific community had gone the other direction in genomics, who would be the heavy that would enforce these rights. Say that Reed Adler had had his way and NIH had in fact gotten these patents, would NIH really go out and sue people who were trying to privatize or tie up these genome sequences? It is not clear yet that that would really have happened, so one of the things that would have to be thought about would be, is there really anybody who would enforce that? I have said enough, and Professor Eisenberg wants to chime in.

*PROFESSOR EISENBERG*:

Yes, I am very happy that you are looking into this. This something I have been thinking about for awhile. I think that the open-source and open genome systems may have more in common ideologically and rhetorically than economically. A few things I wanted to sort of highlight for you, that I think you need to talk about that did not make it into your time-limited presentation. First is the role of the government in all of this, the sponsor of the Human Genome Project, which is just huge. It was very expensive. At least in the

early 90s, this was a lot of money we were talking about. This was really different from open source, where a bunch of guys were working in their garages or whatever, around their workstations. You really need a lot of money in order to crank up these things.

That brings into focus the Bayh-Dole Act.[14] The Bayh-Dole Act quite deliberately constrains the NIH in its ability to stop grantees from pursuing patents. Nonetheless, within that constraint, NIH has tried to play a role as enforcer of open genome. They have tried to do that through their adherence to these international "Bermuda Rules," which you must look into or discuss. These rules, agreed to by the international participants in the Human Genome Project in the mid-90s, in 1995 or 1996, require the deposit of DNA sequence information into public databases within twenty-four hours of getting sequence information. That looked, at the time, to some people, like it was going to short-circuit patent applications because it is tough to get a patent application on file, much less decide what you want to patent, within twenty-four hours. In fact, however, it does not prevent anybody from filing patent applications. It does not prevent others who mine a genetic bank for interesting tidbits of information from filing patent applications. The NIH has not constrained, and I do not think within the law can constrain, grantees from filing patent applications. I am not sure they even want to do so because they want to fulfill the mandate under the Bayh-Dole Act of promoting commercial development of products that come into view as a result of this research. In fact, they have been filing their own patent applications. One final thing I want to say is that your spin on what Reed Adler was up to is one spin –

*PROFESSOR BURK*:

It is Reed's spin. (*laughter*)

*PROFESSOR EISENBERG*:

No, it is not Reed's spin; it is Reed's occasional spin. It is not what Reed would say in writing at the time. Check out Reed's *Science* article on topic, check out Bernadine Healy's *New England Journal of Medicine* article at the time. They were telling a story that it was much more consistent with the Bayh-Dole Act's mandate that they patent the results of their research for the purposes of transferring to the private sector for commercial development. It was not anything like copyleft.

*PROFESSOR BURK*:

Yes, I have actually been very interested in the role of the government also, partly for the reasons that you mentioned, but even more so because of the

[14] 35 U.S.C. §§ 200-212 (2000).

problem that I mentioned of coordination. You have got these two communities that look uncoordinated, yet they are producing these very sophisticated types of outputs. Part of the research that Josh and my colleague David McGowan have been doing in the open-source community indicates that it is not nearly as uncoordinated as it looks. Linus Torvalds keeps a pretty heavy thumb on the development of the Linux kernel using copyleft to make sure that it does not fork and it goes the way he wants it to, and the government, I think, has played a huge role in coordinating, through meetings and transfers and all kinds of other things. It is not nearly as Hayekian as it first appears, and I think you are right that that role is very important.

*PROFESSOR LERNER*:

It is interesting to think about the ideological similarities of these two movements. One of the core issues in both is the hostility toward intellectual property. In the last couple of years, we have seen on the one hand advocates arguing against life forms patenting, against natural products patenting, and against patenting of traditional knowledge. We have also had advocates arguing against business method and software patents. In both, there is an underlying hostility toward intellectual property

But it is not clear that the open source movement must be – or even should be – hostile toward intellectual property. Does not the open-source movement depend at its core on copyright protection, which is then leveraged through licenses to control the manner in which the software is disseminated? Thus, it is using intellectual property to force public disclosure. The approach seems perfectly consistent with patent protection as well. I see no reason why the software underlying the open-source movement should not be unpatentable, so long as the patent rights are licensed parallel to the copyright. Likewise, in the genomic field, the existence of patents on bioinformatic methods, databases, and so forth would not necessarily preclude the sharing of knowledge.

*PROFESSOR BURK*:

You are absolutely right that in the open source area, the copyleft or GNU Public License depends, in the end, on the threat of copyright infringement lawsuit. If you want to work on this piece of code, that is fine. The terms are that you are going to have to make what you do available, and if you do not want to agree to that license and you do the work, then you are the infringer and theoretically we will sue you. I think that is part of the thing that appeals to these people who created the name GNU, Gnu's Not UNIX. This sort of self-referential and recursive use of property rights to try to prevent people from privatizing the material. As I have suggested in the presentation, it seems that that is actually a use of intellectual property very much in line with what Kitch suggested and what Coase suggested to coordinate the development of a resource in a certain way. Now, as we just talked about with Professor Eisenberg, they have not been doing that in the genomic area. They have been

coordinating in other ways through governmental intervention, through grants, through meetings and so on, so it is a very different type of model of how to coordinate that development.

*PROFESSOR O'ROURKE*:

Well, thank you very much, and thank you for coming today.