

# Proprietary versus public domain licensing of software and research products

Alfonso Gambardella<sup>a</sup>, Bronwyn H. Hall<sup>b,c,\*</sup>

<sup>a</sup> *IEGI, Bocconi University, Via Filippetti 9, Milan 20122, Italy*

<sup>b</sup> *Department of Economics, University of California at Berkeley, Berkeley, CA 94720-3880, USA*

<sup>c</sup> *NBER and IFS London*

Available online 30 May 2006

## Abstract

We study the production of knowledge when many researchers or inventors are involved, in a setting where tensions can arise between individual public and private contributions. We first show that, without some kind of coordination, production of the public knowledge good (science or research software or database) is sub-optimal. Then we demonstrate that if ‘lead’ researchers are able to establish a norm of contribution to the public good, a better outcome can be achieved, and we show that the general public license (GPL) used in the provision of open source software is one such mechanism. Our results are then applied to the specific setting where the knowledge being produced is software or a database that will be used by academic researchers and possibly by private firms, using as an example a product familiar to economists, econometric software. We conclude by discussing some of the ways in which pricing can ameliorate the problem of providing these products to academic researchers.

© 2006 Published by Elsevier B.V.

*Keywords:* Open source; Software; Intellectual property; Scientific research; Databases

## 1. Introduction

In the modern academic research setting, many disciplines produce software and databases as a by-product of their own activities and also use the software and data generated by others. As Dalle (2003) and Maurer (2002) have documented, many of these research products are distributed and transferred to others using institutions that range from commercial exploitation to ‘free’ forms of open source. Many of the structures used in the latter case resemble the traditional ways in which the ‘Republic of Science’ has ensured that research spillovers are available at low cost to all. But in some cases, moves

toward closing the source code and commercial development take place, often resulting either in the disappearance of open source versions or in ‘forking’, where an open source solution survives simultaneously with the provision of a closed commercial version of the same product. This has also created tensions between the reward systems of the ‘Republic of Science’ and the private sector, especially when the production of research software or the creation of scientific databases is carried out in academic and scientific research environments (see also Hall, 2004).

As these inputs to scientific research have become more essential and their value has grown, a number of questions and problems have arisen surrounding their provision. How do we ensure that incentives are in place to encourage their supply? How does market and non-market production of these knowledge inputs interact? In this paper, we address some of these questions. We

\* Corresponding author. Tel.: +1 510 642 0822;

fax: +1 510 642 0638.

*E-mail address:* [bhhall@econ.berkeley.edu](mailto:bhhall@econ.berkeley.edu) (B.H. Hall).

develop a framework that highlights the difficulties of sustaining the production of knowledge when it is the outcome of a collective enterprise. Since the lack of coordination among the individual knowledge producers is typically seen as the culprit for the under-provision of public knowledge, the latter can be sustained by institutional devices that encourage such coordination. A key idea of the paper is that the general public license (GPL) used in the provision of open source software is one of such mechanisms. We then discuss another limitation in the production of this type of knowledge. To make it useful for commercial or other goals, one needs complementary investments (e.g. development costs). If the knowledge is freely available, there could be too many potential producers of such investments, which reduce the incentives of all of them to make the investments in the first place. Paradoxically, if the knowledge were protected, its access would be more costly, which may produce the necessary rents to enhance the complementary investments. Yet protecting upstream knowledge has many drawbacks, and we argue that a more effective solution is to protect the downstream industry products. Finally, we discuss how our framework and predictions apply to the provision of scientific software and databases.

An example of the difference between free and commercial software solutions that should be familiar to most economists and scientific researchers is the scientific typesetting and word-processing package *TeX*.<sup>1</sup> This system and its associated set of fonts was originally the elegant invention of the Stanford computer scientist, Donald Knuth, also famous as the author of the *Art of Computer Programming*, the first volume of which was published in 1969. Initially available on mainframes, and now widely distributed on UNIX and personal computer systems, *TeX* facilitated the creation of complex mathematical formulas in a word-processed manuscript and the subsequent production of typeset camera-ready output. It uses a set of text-based computer commands to accomplish this task rather than enabling users to enter their equations via the graphical WYSIWYG interface now familiar on the personal computer.<sup>2</sup> Although straightforward in concept, the command language is complex and not easily learned, especially if the user does not use it on a regular basis. And although many academic users

still write in raw *TeX* in spite of the fact that they work on a system with a graphical interface such as *Windows*, there now exists a commercial program, *Scientific Word*, which provides a WYSIWYG environment for generating *TeX* documents, albeit at a considerable price when compared to the freely distributed original.

This example illustrates several features of the academic provision of software that we will discuss in this paper. First, it shows that there is willingness to pay for ease of software use even in the academic world and even if the software itself can be obtained for free. Second, the most common way in which software and databases are supplied to the academic market is a kind of hybrid between academic and commercial, where they are sold in a price-discriminatory way that preserves access for the majority of scientific users. Such products often begin as open source projects directed by a ‘lead’ user, because the culture of open science is quite strong in the developers and participants. Nevertheless, they are eventually forced into the private sector as the market grows and non-developer users demand support, documentation, and enhancements to the ease of use.

In the next section we discuss some basic aspects of the problem of creating incentives for the production of knowledge when many producers are involved. Section 3 discusses our analytic framework which shows that without some kind of coordination, production of the public knowledge good (science or research software or database) is sub-optimal, and that the GPL can solve the problem at least in part. Section 4 focuses on complementary investments. Sections 5 and 6 apply our framework to the specific setting where the knowledge being produced is software or a database that will be used by academic researchers and possibly also by private firms, using as an example a product familiar to economists, econometric software. We conclude by discussing some of the ways in which pricing can ameliorate the problem of providing these products to academic researchers. Appendix A develops the technical details of our model in Section 3.

## 2. Incentives for knowledge production with many producers

The design of incentive systems that reward inventors and knowledge producers and encourage dissemination of their output has been a familiar issue to economists and other scholars for a long time (e.g. Nelson, 1959; Arrow, 1962; Scotchmer, 1991). If anything, the issue has become more important today with the advent of the Internet and other computer networking methods. The principal effect of the increase in computer networking

<sup>1</sup> This brief history of *TeX* is drawn from the *TeX* User’s Group website, <http://www.tug.org> (TeX, 2005). In giving a simplified overview, we have omitted the role played by useful programs based on *TeX* such as *LaTeX*, etc. (see the website for more information).

<sup>2</sup> WYSIWYG is a widely used acronym in computer programming design that stands for “what you see is what you get”.

and Internet use is that it lowers the marginal cost of distributing codified knowledge to the point where it is essentially zero. This in turn has the potential to reduce incentives for production of such knowledge or to increase the demands of the producers for protection of their property rights to the knowledge. Hence there is a felt need for undertaking additional efforts to understand the production of knowledge, and for thinking about new approaches to policy.

To address these issues, we must first ask what motivates the producers of knowledge. Key factors identified in the literature are curiosity and a taste for science, money, the desire for fame and reputation, and as a secondary goal, promotion or tenure (Stephan, 1996). The latter two goals are usually achieved via *priority* in publication, that is, being the first to get a discovery into print. Although monetary income is clearly a partial motivation in the search for reputation and promotion, considerable evidence exists that for researchers in universities and public research organizations with some level of guaranteed income, the first motive – intellectual curiosity – is of overriding importance (e.g. Isabelle, 2004). For this type of researcher, the desire for financial rewards is often driven by the desire to fund their own scientific research (Lee, 2000) rather than by consumption *per se*. Scientists' motivations also are coloured by the culture in which they are embedded, with traditional norms giving way to a more market-oriented view among some younger scientists today (Isabelle, 2004; Owen-Smith and Powell, 2001).

Several scholars (e.g. Merton, 1957, 1968; David, 1993) have described the two regimes that allocate resources for the creation of new knowledge: one is the system of granting intellectual property rights, as exemplified by modern patent and copyright systems, the other is the 'open science' regime, as often found in the realm of 'pure' scientific research and sometimes in realm of commercial technological innovation, often in infant industries (Allen, 1983). Today we also see this system to a certain extent in the production of free and open source software. The first system assigns clear property rights to newly created knowledge that allow the exclusion of others from using that knowledge, as well as the trading and licensing of the knowledge. As is well known, such a system provides powerful incentives for the creation of knowledge, at the cost of creating temporary monopolies that will tend to restrict output and raise price. Additionally, in such systems, the transaction costs of combining pieces of knowledge or building on another's knowledge may be rather high, and in some cases achieving first or even second best incentives via *ex post* licensing may be impossible (Scotchmer, 1991). The use of

other firms' knowledge output will often require payment or reciprocal cross-licensing, which means negotiation costs have to be incurred. Finally, obtaining IP rights usually requires publication, but only of codified knowledge, and trade secrecy protection is often used in addition.

The second set of institutional arrangements, sometimes referred to as the norms governing the 'Republic of Science', generates incentives and rewards indirectly; the creation of new knowledge is rewarded by increased reputation, further access to research resources, and possible subsequent financial returns in the form of increased salary, prizes, and the like (Merton, 1957, 1968). This system relies to some extent on the fact that individuals often invent or create for non-pecuniary reasons like curiosity. Dissemination of research results and knowledge is achieved at relatively low cost, because assigning the 'moral rights' to the first publisher of an addition to the body of knowledge gives creators an incentive to disseminate rapidly and broadly. Therefore, in this system the use of others' output is encouraged and relatively cheap, with the cost being appropriate citation and possibly some reciprocity in sharing knowledge. However, it is evident that this system cannot capture the same level of private economic returns for the creation of knowledge. Inventors must either donate their work or receive compensation as clients of public or private patrons.<sup>3</sup>

Hall (2004) highlights the tension that arises when these two systems come up against each other. For example, it is common for the difference in norms and lack of understanding of the potential partner's needs and goals to produce breakdowns in negotiations between industry and academe. These breakdowns can have an economic as well as cultural cause, as shown by Anton and Yao (2002) in a study of contracting under asymmetric information about the value of the knowledge to be exchanged. In addition, there is the simple fact that both systems rely on reciprocal behaviour between both parties to a knowledge exchange, so that contracting between participants in the two difference systems becomes subject to misunderstanding or worse. This is illustrated by the reaction of the genomic industry in the U.S. when asked to take out licenses to university-generated technology; once the university starts acting like a private sector firm, there is a temptation to start charging them for the use of the outputs of industry

<sup>3</sup> We can subsume both cases as instances of 'patronage' — self patronage of the donated efforts is a special case of this (see David, 1993; Dasgupta and David, 1994).

research, with consequent negative effects on researchers who still believed themselves part of the ‘open science’ regime.

In fact, notice should also be taken of an important variation of the ‘open science’ regime for the sharing of knowledge production outputs, one which has existed many times in the development of industry throughout history: the free exchange and spillover of knowledge via personnel contact and movement, as well as reverse engineering, without resort to intellectual property protection. This has become known as the system for ‘collective invention’. Examples include the collective invention in the steel and iron industry described by Allen (1983) (see also Von Hippel, 1987), the development of the semiconductor industry in Silicon Valley (Hall and Ziedonis, 2001), the silk industry in Lyons during the *ancien régime* that was described by Foray and Hilaire-Perez (2005), and the collective activities of communities of users who freely distribute information to the manufacturers (Harhoff et al., 2003). In these environments, most of which are geographically localized innovation areas with social as well as business relationships that build trust (or at least knowledge of whom to trust), the incentive system for the production and *exchange* of knowledge is somewhat different than in either of the other two systems.

The first and most obvious difference is that the production of ‘research’ in the industry setting is supported not by public or private patronage but by commercial firms that finance it by the sale of end-product that incorporate their discovery. Because rewards come from the sale of products rather than information itself, as they do in the conventional IP-based system, the sharing of information about incremental innovations is motivated by different considerations than in the case of the OS regime. Although priority is not per se valuable except in the sense that it may confer lead-time for production, shared knowledge, especially about incremental improvements to a complex product, is perceived to be useful and essential for the progress of the entire industry, including the firm that shares the knowledge. When an industry is advancing and growing rapidly, the desire to exclude competitors from the marketplace is not as strong as when an industry reaches maturity. An implication is that this form of free exchange of knowledge tends to collapse, or is unstable over time, as has happened in many of the historical examples. In the next section we will try to capture this idea and discuss some conditions under which the academic or industry-based OS regime might break down.

### 3. ‘Public domain’ vs. ‘proprietary’ research

#### 3.1. Configuration of the open source equilibrium

When do the different systems of knowledge generation and sharing discussed in the previous section develop, and when they might be expected to break down? In this section we address these questions. To make our argument more precise, we provide a simple formalization in Appendix A. Below we discuss the intuitions and the implications of our model.

As discussed in the previous section, many researchers face a trade-off. They can put a given research outcome in the public domain or seek private profits from it. As a stylized representation, in the former case they enjoy no economic rents, while in the latter they restrict public diffusion of their findings, seek property rights on them, and gain monetary income. We label the first mode as ‘public domain’ (PD), and the second as ‘proprietary research’ (PR). As also noted earlier, this framework encompasses many situations, such as academic scientists who could publish their research findings vis-à-vis holding patents or other property rights on them (Dasgupta and David, 1994); software developers who contribute to open source software as opposed to patenting their programs (Lerner and Tirole, 2002); user-inventors who transfer their inventions to the producers rather than protecting them as intellectual property and then selling them (Von Hippel, 1988; Von Hippel and von Krogh, 2003; Harhoff et al., 2003); and communities of technologists who coordinate to share their ‘collective’ inventions, as opposed to keeping their knowledge secret (Allen, 1983; Nuvolari, 2004; Foray and Hilaire-Perez, 2005).

Like any individuals, researchers gain utility from monetary income, but their utility also increases with the stock of public domain (PD) knowledge. Their benefits from this knowledge are from two sources: their own contributions and other contributions. First, they enjoy utility from the fact that they contribute to public knowledge. This is because they ‘like’ contributing to PD knowledge per se, or because they enjoy utility from a larger stock of public knowledge and hence they wish to contribute to its increase. There could also be instrumental reasons. Contribution to public knowledge makes their research visible, which provides fame, glory or potential future monetary incomes in the form of increased salary, funding for their research, or consultancy. Second, the researchers gain utility from the fact that others contribute to PD knowledge. Again, this could be because they care about the state of public knowledge. In addition, a greater stock of public knowledge provides

a larger basis for their own research, which implies that, other things being equal, they would like others contribute to it.

We assume that the benefits from the contributions of other researchers to public knowledge will be enjoyed whether one works under PD or in the proprietary research regime (PR). This implies that a researcher will operate under PD if the benefit that she enjoys from her public contribution is higher than the foregone monetary income from not privatizing her findings. In [Appendix A](#) we show that in equilibrium this is true of all the researchers that operate under PD, while the opposite is true of the researchers that operate under PR. In general, the equilibrium will involve a share of researchers operating under PD or PR that is between 0 and 1. The first prediction of our analysis is, then, that the two regimes can co-exist, as we shall also see with some examples in the following sections.

Our model also predicts that new profit opportunities common to all the researchers in a field reduce the share of PD researchers in equilibrium, while a stronger taste for research (e.g. because of particular systems of academic values) raises it. There is fairly widespread evidence that in fields like software or biotechnology there are pressures on academic researchers to place their findings in a proprietary regime. Also, our examples in the later sections show that shifts from academic to commercial software are more prominent when the market demand for the products increase, which raises the profitability of the programming efforts. Finally, there are several accounts of the fact that tension between industrial research and academic norms become higher if university access to IPRs is increased ([Cohen et al., 1998](#); [Hall et al., 2001](#); [Hertzfeld et al., this issue](#)). As these authors report, such tension has already been observed in the US, as the latter country has pioneered the trend towards stronger IPRs and the use of intellectual property protection by universities, but it is becoming more pronounced in Europe as well, as European universities follow the path opened up by the US system ([Geuna and Nesta, this issue](#)). [Collins and Wakoh \(1999\)](#) describe similar changes in Japan and describe how the regime shift to patenting by universities is inconsistent with the previous system of collaborative research with industry in that country, implying increasing stress for the system.

### 3.2. *Instability of open source production*

Our model also shows that the only way to get a *stable* equilibrium configuration with individuals operating under open sharing rules is when there is coordination

among them. Otherwise, the sharing (cooperative) equilibrium tends to break down because some individuals find it in their interest to defect. The instability of the open sharing equilibrium is just an application of the famous principle by [Olson \(1971\)](#) that without coordination the collective action is hard to sustain. Our contribution is simply to highlight that Olson's insight can be applied to the analysis of the instability of open systems. When many researchers contribute to PD knowledge, an individual deviation to PR is typically negligible compared to the (discrete) jump in income offered by proprietary regime. Thus, individually, the researchers always have an incentive to deviate.

Another way to see this point is that some of the tensions that are created in open research systems can be attributed to the asymmetry between the open and the proprietary modes. The researchers shift to proprietary research only if individually profitable. By contrast, in the collective production of knowledge, a desirable individual outcome depends on the actions of others. In our framework, this is because the individuals care about the fact that others contribute to the stock of knowledge, and because this may affect their benefits from their own contribution as well. As we show in [Appendix A](#), this creates situations in which the lack of coordination produces individual incentives to deviate, in spite of the fact that collectively the researchers would like to produce under PD. The intuition is that a group of individuals can produce a sizable increase in the stock of public knowledge if they jointly deviate from the PR regime. Thus, if there were commitment among them to stay with the PD rules, they could be better off than with private profits. In turn, this is because the larger the group of people who deviate in a coordinated fashion, the higher the impact on the public knowledge good, while the private profits, which do not depend as much on the collective action, are not affected substantially by the joint movement of researchers from one regime to the other. But even if they all prefer to stay with the PD system, because of the larger impact of their PD contributions as a group, individually they have an incentive to deviate because, if the others stay with PD, the individual deviation does not subtract that much from public knowledge, while it does produce a discrete jump in the individual's private income. Since everyone knows that everyone else faces this tension, and could deviate, it will be difficult to keep the researchers under the PD system unless some explicit coordination or other mechanism is in place.

Ultimately, this asymmetry in the stability of the two configurations suggests why there may be a tendency to move from public to private production of knowledge, while it is much harder to move back from private to



public. The implication is that there is little need for policy if more proprietary research is desirable, as the latter is likely to arise naturally from the individual actions. By contrast, policy or institutional devices that could sustain the right amount of coordination is crucial if the system under-invests in knowledge that is placed in the public domain.

### 3.3. Generalized public license ('copyleft') as a coordination device

The generalized public license (GPL) used in open source software can be an effective mechanism for obtaining the required coordination. As discussed by [Lerner and Tirole \(2002\)](#) *inter alia*, with a GPL the producer of an open source program requires that all modifications and improvements of the program are subject to the same rules of openness, most notably the source code of all the modifications ought to be made publicly available like the original program.<sup>4</sup> To see how a GPL provides the coordination to solve the Olson problem, imagine the following situation. There is one researcher who considers whether to launch a new project or not. We call her the 'originator'. She knows that if she launches the project, others may follow with additional contributions. The latter are the 'contributors'. If the originator attaches a GPL to the project, the contributors can only join under PD. If no GPL is attached, they have the option to privatize their contribution. Of course, once (and if) the project is launched, the contributors always have the option not to join the project and to work on some alternative activities. Given the expected behaviour of the contributors, the originator will choose whether to launch the project or not. She also has potential alternatives. If she decides to launch it, she will choose whether to put her contribution under PD or PR, and if the former she considers whether to attach a GPL to the project. We can safely rule out the possibility that the originator operates under PR and attaches a GPL to the project. It would be odd to think that she can enforce open source behaviour given that she does not abide by the same rules.

The key implication of a GPL is that it increases the number of contributors operating under PD. The intuition, which we formalize in [Appendix A](#), is simple. Without a GPL the contributors have three choices: to

work on the project under PD or under PR, or not to join because they have better alternatives. PD contributors to the project will still choose PD if a GPL is imposed. If they preferred PD over both PR and other alternatives, they will still prefer PD if the PR option is ruled out. Those who did not join the project will not join with a GPL either. They preferred their alternatives over PD and PR, and will still prefer them if PR is not an option. Finally, some of those who joined under PR will join under PD instead, while others who joined under PR will no longer join the project. As a result, a GPL reduces the total number of researchers who join the project, but raises the number of researchers working under PD. The reduced number of participants is consistent with the fact that the GPL is a restriction on the behaviour of the researchers. However, this is a small cost to the public diffusion of knowledge because those who no longer participate would have not joined under PD. By contrast, the GPL encourages some researchers who would not have published their results without the GPL to do so.

Given the behaviour of the contributors, will the originator launch the project and issue a GPL? We know that the originator, like any other researcher, enjoys greater utility from a larger size of the public knowledge stock. At the same time, she enjoys utility from monetary income or, as we noted, from alternative projects. Here we want to compare her choice when she can employ a GPL vis-à-vis a world in which there is no GPL. With a GPL she knows that the number of contributors to public knowledge increases, which in turn increases the size of the expected public knowledge stock when compared to a no-GPL case. As a result, when choosing whether to launch the project under PD with a GPL, under PD and no GPL, under PR, or to work on alternative projects, she knows that the GPL choice raises the future public knowledge stock in the area while not raising her monetary income from the project or her utility from alternatives. This makes it more likely that the originator will choose to work on the project under PD cum GPL. More generally, a GPL will increase the number of projects launched under PD and the size of the public knowledge contributions.

To summarize, the way the GPL works is by giving rise to an implicit coordination among a larger number of researchers to work on PD. The originator knows that there will be researchers who would prefer PR but choose PD if the former opportunity is not available, while all those who would choose PD will stick to it in any case. This enlarges the number of expected PD researchers, thereby placing greater advantages on the PD choice. Our intuition is that those with a strong taste for PD research will always work under PD, whether there is a

<sup>4</sup> There are many variants of a GPL, with different possibilities of privatizing future contributions (see, for example, [Lerner and Tirole, 2005](#)). However, in this paper we want to focus on some broad features of the effect of a GPL as a coordinating device, and therefore we simply consider the extreme case in which the GPL prevents any privatization of the future contributions.

GPL or not. By contrast, those with a high opportunity cost will never join the project. But those who have a small opportunity cost, and a weak taste for PD research might contribute via PD if a GPL is introduced. The GPL then lures people who are on the border between doing PD research on the project or not. For example, a GPL may be crucial to enhance the participation under PD of young researchers, who do not have significant opportunity costs (e.g. because they do not yet have high external visibility), but who do not have a strong taste for PD research either, and hence would privatize their findings if profitable to do so. There might also be dynamic implications — e.g. the GPL helps young researchers to ‘acquire’ a taste for PD research. This might help create a system of norms and values for public research that could sustain the collective action. We leave a more thorough assessment of such dynamic implications to future research.

### 3.4. *Nature and consequences of the GPL coordination*

A GPL is most effective as a coordination device when the opportunity cost of the individual researchers and the private profits from contributing to the project, are not positively correlated. Suppose that they were. This could arise because there is some common element between the two factors. For example, an individual researcher could be effective in commercializing knowledge in any field because he belongs to institutions (university or other) that encourage the commercialization of knowledge. In this case, the contributors to the project, who have low opportunity costs, also have low private profits from contributing to the project via PR. A GPL would not make a big difference because a very large fraction of the contributors to the project will do so under PD since their private rewards are low in any case. Hence, a GPL induces few researchers to switch from PR to PD. In turn, this has a small effect on the choice of the originator to launch the project under PD vis-à-vis PR because the number of additional PD contributors with a GPL is small. By contrast, if they are not positively correlated, some of the contributors to the project, who have low opportunity costs, will have high private rewards from PR. They could be encouraged by a GPL to switch to PD. As a result, the number of PD contributors could be sizably different with a GPL, with implied greater opportunities for PD rather than PR research.

The independence between the opportunity cost and the private rewards, as opposed to positive correlation, may be associated with the novelty of the project.

When the projects are in new areas, the opportunities of the individuals may change substantially, and the researchers who might profit the most from the new projects can be different from those who benefited in the old projects. New skills, or new forms of learning are necessary in the new fields, and the people who have made substantial investments in the old projects may have greater difficulties in the new areas (see, for example, [Levinthal and March, 1993](#)). In these cases, researchers with low opportunity costs may instead find that they have great opportunities to commercialize knowledge in the new fields (high private rewards). Thus, the GPL is more likely to be a useful coordination device when the project is in a new field rather than an incrementally different one from previous projects, and when it is socially desirable to run these projects under PD.

Our mechanism relies on the fact that there is enforcement of the GPL. But can the ‘copyleft’ system be enforced? In some settings people seem to abide by the copyleft rules, as [Lerner and Tirole \(2002\)](#) have noted, in spite of the lack of legal enforcement. In many situations, there may be a reputation effect involved when the copyleft agreement is not complied with. In this respect, the reason why a copyleft license may be useful is that without it, it may not be clear to the additional contributors whether the intention of the initial developers of the project was to keep it under PD or not. But if the intention is made explicit, deviations may be seen as an obvious and explicit challenge to the social norms, and this may be sanctioned by the community. The GPL then acts as a signal that clears any potential ambiguities about individual behaviour and respect for the social norms. Even in science, if a researcher developed a certain result, others may build on it, and privatize their contributions. This might be seen as a deviation from the social norms. While this behaviour could be sanctioned, according to the strength with which the norms of open science are embedded in and pursued by the community, with no explicit indication that the original contributor did not want future results from her discoveries to be used for private purposes, the justification for the sanctions or the need for them can be more ambiguous.

A GPL removes ambiguity about the original intentions of the developers, and any behaviour that contradicts the GPL is more clearly seen as not proper. This reduces privatization of future contributions compared to a situation with no GPL, increases the expectations that more researchers will make their knowledge public, and, other things being equal, creates greater incentives to make projects public in the first place. It is in this respect that we think that explicit indications of the norms may

be a stronger signal than the mere reliance on the unwritten norms of open science or open source software.

A related point is that the literature has typically been concerned with the need to protect the private property of knowledge when this is necessary to enhance the incentives to innovate. The inherent assumption is that when it is not privately protected, the knowledge is by default public, and it enriches the public domain. Yet, our model points out that this is not really true. The public nature of knowledge needs itself to be protected when commitments to the production of knowledge in the public domain is socially desirable. In other words, there is a need for making it explicit that the knowledge has to remain public, and this calls for positive actions and institutions to protect it. Not allowing for private property rights on some body of knowledge is not equivalent to assuming that the knowledge will be in the public domain. One may then need to assign property rights not just to private agents but also to the public. For example, the IPRs are typically thought of as being property rights to private agents. Yet we also need to have institutions that preserve the public character of knowledge. The copyleft license is a beautiful example of this institutional device. A natural policy suggestion is therefore to make it legal and enforceable as copyright, patents and other private-based IPRs.

#### 4. Complementary investments in open source production

Another feature of traditional open source or academic software production that we alluded to in the introduction is that it normally requires additional investments that enhance the usefulness and value of the scattered individual contributions, or it simply requires investments to combine them. For example, while several individuals can contribute to the development of a whole body of scientific knowledge, there must be some stage in which the ‘pieces’ are combined into useful products, systems or transferable knowledge. Some scientists or most likely some specialized agents, i.e. academic licensing offices or firms, normally perform this function. A typical example is when scientific knowledge needs substantial downstream investments to become economically useful technologies or commercializable products. *Thursby et al. (2001)* reported that this is often the case for university research outputs. The latter activities are normally performed by firms. In software, additional investments are often required to enhance the usability of the software for those who did not develop it, and to produce documentation and support. The need

for additional investments in open source production, or more generally in tasks that rely on public domain knowledge, has some specific implications that we want to discuss in this section.

The problem is that the (downstream) ‘assembling’ agent needs some profits in order to carry out the investments that are necessary to produce the complementary downstream assets of the good. Since the downstream assembling agents are typically firms, we now refer to them as the latter. There are two issues. First, the firm needs to obtain some economic returns to finance its investment. Clearly, there are many ways to moderate its potential monopoly power so that the magnitude of the rents will be sufficient to make the necessary investments but not high enough to produce serious extra-normal profits. However, it would be difficult for the firm to obtain such rents if it operated under perfect competition, or if it operated under an open, public domain system itself.

The second issue is more subtle. The firm uses the public domain contributions of the individual agents (software programmers, scientists, etc.) as inputs in its production process. If these contributions are freely available in the public domain, and particularly if they are not available on an exclusive basis, many downstream firms can make use of them. As a result, the downstream production can easily become a free-entry, perfectly competitive world, with many firms having access to the widely available knowledge inputs. If so, each firm could not earn enough rents to carry out the complementary investments. This would be even harder for the individual knowledge producers who are normally scattered and have no resources to cover the fixed set-up costs for the downstream investments. The final implication is that the downstream investments will not be undertaken, or they will be insufficient. Of course, there can be other factors that would provide the firms with barriers to entry, thereby ensuring that they can enjoy some rents to make their investments. However, in productions where the knowledge inputs are crucial (e.g. software), the inability to use them somewhat exclusively can generate enough threats of widespread entry and excessive competition to discourage the complementary investments.

Paradoxically, if the knowledge inputs were produced under proprietary rules, the producers of them could charge monopoly prices (e.g. because they could obtain an exclusive license) or at least enjoy some positive price cost margins. This raises the costs of the inputs. In turn, this heightens barriers to entry in the downstream sector, and adjusts the level of downstream investment upward. In other words, if the inputs are



freely available, there could be excessive downstream competition, which may limit the complementary investments. If they are offered under proprietary rules, the costs of acquiring the inputs are higher, which curbs entry and competition, and allows the downstream firms to earn enough rents to carry out such investments.<sup>5</sup>

However, the privatization of the upstream inputs has several limitations. For one, as Heller and Eisenberg (1998) have noted, the complementarity among the ‘pieces’ of upstream knowledge produced by the different individuals can give rise to the so-called problem of the ‘anti-commons’. That is, after all the other rights have been collected under a unique proprietorship, the final owner of a set of complementary inputs can enjoy enormous monopoly power. This is because by withholding his own contribution, he can forestall the realization of the whole technology, especially when the complementarity is so tight that each individual contribution is crucial to make the whole system work. The possibility of ex post hold-up can discourage the effort to collect all the complementary rights ex ante, and therefore prevent the development of the technology. Another limitation of the privatization of the upstream inputs is the one discussed in the previous section. With copyleft agreements, more people can contribute to the public good. The decentralized nature of the process by which scientists or open source software producers operate has typically implied that the network of public contributors to a given field can be so large that the overall improvements can be higher than can be obtained within individual organizations, including quite large ones. Some evidence that open source projects also increase the quality of software output has been supplied by Kuan (2002).

One solution to the problem of paying for complementary downstream investment is allowing for property rights, and particularly intellectual property rights, on the innovations of the downstream producer. This would, of course, raise its monopoly power and therefore curb excessive competition. At the same time, it avoids attaching IPRs to pieces of upstream knowledge, thereby giving rise to the problems of the anti-commons, or to reduced quality of the upstream knowledge. In addition, the downstream producer would enjoy rights on features of the innovation that are closer to his own real contribution to the project, that is the development of specific

downstream investments. Clearly, this also implies that the IPRs thus offered are likely to be narrower, as they apply to downstream innovations as opposed to potentially general pieces of knowledge upstream. At the same time, they are not likely to be as narrow as in the case of small individual contributions to an open software module or a minor contribution to a scientific field, which can give rise to the fragmentation and hold-up problems discussed earlier.

## 5. Academic software and databases

In this section we draw some implications for the provision of scientific software and databases from the model and discussion in the previous two sections, and then go on to discuss the possible modes in which they could be provided. First, this type of activity is more likely to be privatized than scientific research itself because there is greater and more focused market demand for the product, because norms are weaker due to weaker reputation effects, and because there are more potential users who are not inventors (and who do not participate in the production of the good). Second, there could easily be both public and private provision at the same time, because such an equilibrium can be sustained when there are different communities of researchers with different norms. Third, as the market for a particular product grows, privatization is likely simply because the individual’s discrete return to privatization has increased. Finally, when the components to a valuable good are produced under public domain rules, free entry in the downstream industry producing a final good based on those components implies too few profits for those undertaking investments that will enhance the value of the good. The final producers have to earn some rents to be able to make improvements beyond the mere availability of research inputs.

The privatization of scientific databases and software has both advantages and disadvantages. With respect to the latter, David (2002) has emphasized the negative consequences of the privatization of scientific and technical data and information. One of the most important drawbacks is the increase in cost, sometimes substantial, to other scientists, researchers or software developers for use of the data in ways that might considerably enhance public domain knowledge. A second is that the value of such databases for scientific research is frequently enhanced by combining them or using them in their entirety for large-scale statistical analysis, both of which activities are frequently limited when they are

<sup>5</sup> This argument should be familiar as it is the same as the argument used by some to justify the Bayh-Dole Act in the U.S. and the granting of exclusive licenses for development by universities.

commercially provided.<sup>6</sup> Maurer (2002) gives a number of examples of privatized databases that have somewhat restricted access for academic researchers via their pricing structure or limitations on reuse of the data (such as Swiss-PROT, Space Imaging Corporation, Incyte, and Celera). In a paper originally intended to form part of this special issue, David (2000) cites the case of the privatization of Landsat images under the Reagan administration, which led to a 10-fold increase in the price of an image. In terms of our model, the potential to privatize scientific and technical data and information implies that a smaller number of researchers will contribute to the public good and a smaller stock of public knowledge will be produced, thus frustrating the launch of projects undertaken under public diffusion rules.

At the same time, a common argument in favour of the privatization of databases is that it helps in the development of a database-producing industry, and more generally of an industry that employs these data as inputs. A similar argument can be used more broadly for software. For example, the recent European Directive that defines the terms for the patenting of software in Europe (European Commission, 2002) was largely justified by the argument that it would encourage the formation of a software industry in niches and specialized fields. Although it is sometimes true that exclusivity can have positive effects on the provision of information products, it is also true that there can be drawbacks like those suggested earlier (e.g. fragmentation of IPRs, little contribution to public domain knowledge, restricted access when welfare would be enhanced with unlimited access) to the privatization of knowledge inputs. At times, one can obtain similar advantages by allowing for the privatization of the outputs that can be generated using the database or software in question. For example, the discovery of a useful application associated with a particular gene that is obtained by use of a genomic database is patentable in most countries. Similarly, in the case of the econometric software example used later in the paper, consulting firms such as Data Resources Inc. or Chase Econometrics marketed the results of estimating econometric models using software whose origins were in the public domain. Following our earlier argument, by allowing for the privatization of the downstream output we enable the industry to obtain enough rents to make the

necessary complementary investments, while avoiding the limitations of privatizations in the upstream knowledge.

There are, however, limits to this particular strategy for ensuring that scientific databases and software remain in the public domain while downstream industries based on these freely available discoveries can earn enough profit to cover their necessary investments. The difficulty, of course, is that in the case of generally useful information products, a firm selling a particular product, the input for which is an upstream academic product, has no reason to undertake the enhancements to the upstream product that would make it useful to others, unless the firm can sell the enhanced product in the marketplace. Yet this is what we were trying to avoid, and what is ruled out by a GPL.

In fact, we now turn to a discussion of an alternative way in which such goods can be provided. The production of information products including software and databases has always been characterized by large fixed costs relative to marginal costs, but the cost disparity has grown since the advent of the Internet. In practice, the only real marginal costs of distribution arise from two sources: the support offered to individual users (which in many cases has been converted into a fixed cost by requiring users to browse knowledge bases on the web); and the congestion costs that can occur on web servers if demand is too great.<sup>7</sup> Standard economic theory tells us that when the production function for a good is characterized by high fixed costs and low marginal costs, higher welfare can often be achieved by using discriminatory pricing, charging those with high willingness to pay more in order to offer the good to others at lower prices, thus increasing the overall quantity supplied. The problem with applying this mechanism generally is the difficulty of segmenting the markets successfully and of preventing resale.

In the case of academic software and databases, however, it is quite common for successful price-discriminating strategies to be pursued.<sup>8</sup> There are sev-

<sup>6</sup> The usual commercial web-based provision of data is based on a model where the user constructs queries to access individual items in the database, like looking up a single word in the dictionary. The pricing of such access reflects this design and is ill-suited (i.e. very costly) for researcher use in the case where research involves studying the overall structure of the data.

<sup>7</sup> This can be a real cost. The U.S. Patent Office, which provides a large patent database free to the public at large on its web server, has a notice prominently posted on the website saying that use of automated scripts to access large amounts of this data is prohibited and will be shut down, because of the negative impact this has on other individuals making queries.

<sup>8</sup> Another type of academic information product deserves mention here, academic journals. The private sector producers of these journals face the same type of cost structure and have pursued a price discrimination strategy for many years, discriminating between library and personal use, and also according to the income levels of purchasers in some cases, where income level is proxied by country of origin.

eral reasons for this: (1) segmentation is fairly easy because academics can be identified via addresses and institutional web information; (2) resale is difficult in the case of an information product that requires signing on to use it and also probably not very profitable; (3) the two markets (academic and commercial) have rather different tastes and attitudes toward technical support (especially towards the speed with which it is provided) so the necessary price discrimination is partly cost-based.

## 6. Case study: econometric software packages

As an illustration of the pattern of software development in the academic arena, we present some evidence about a type of product familiar to economists that has largely been developed in a university research environment but is now widely available from commercial firms: packaged econometric software. Our data are drawn primarily from the excellent surveys on the topic by Renfro (2003, 2004). We have supplemented it in places from the personal experience of one of the co-authors, who participated in the activity almost from its inception. The evidence supplied here can be considered illustrative rather than a formal statistical test of our model, since the sample is relatively small. To form a complete picture of the phenomenon of software and database commercialization in academia, it would be necessary to augment our study with other case studies. (For example, see Maurer, 2002 for a good review of methods of database provision in scientific research.)

Econometric software is very much a by-product of empirical economic research activity, which is conducted largely at universities and non-profit research institutions and to a lesser extent in the research departments of banks and brokerage houses. It is an essential tool for the implementation of statistical methods developed by econometric theorists, at least if these methods are to be used by more than a very few specialists. To a great extent, this type of software originated during the 1960s, when economists began to use computers rather than calculating machines for estimation, and for the first time had access to more data than could comfortably be manipulated by hand. A typical software package is implemented using a simple command language, enabling the use of a variety of modelling, estimating and forecasting methods on datasets of varying magnitudes. Most of these packages are now available for use on personal computers, although their origins were often a mainframe computer implementation. (For a complete history of the development of this software, see Renfro, 2003.)

Like most software, econometric software can be protected via various IP measures. The most important is a combination of copyright (for the specific implementation in source code of the methods provided) and trade secrecy (whereby only the ‘object’ code, or machine-readable version of the code, is released to the public). This combination of IP protection has always been available but has only become widely used during the personal computer era. Prior to that time, distributors of academic software usually provided some form of copyrighted source code for local installation on mainframes, and relied on the fact that acquisition and maintenance were performed by institutions rather than a single individual to protect the code. This meant the source code could be modified for local use, but because the size of the potential market for ‘bootleg’ copies of the source was rather small, piracy posed no serious competitive threat. The advent of the personal computer, which meant that in many cases software was being supplied to individuals rather than institutions changed this situation, and today the copyright-trade secrecy model is paramount.<sup>9</sup> Thus, it is possible to argue that developments in computing have made the available IP protection in the academic software sector stronger at the same time that the potential market size grew, which our model implies will lead to more defection from public domain to proprietary rules.

In Table 1, we show some statistics for the 30 packages identified by Renfro. The majority (20 of the 30) have their origins in academic research, either supported by grants or, in many cases, written as a by-product of thesis research in a student’s own time.<sup>10</sup> A further five were written specifically to support the modelling or research activities of a quasi-governmental organization such as a central bank. Only five were written with a specific commercial purpose in mind. Two of those five were forks of public domain programs, and in contrast to those of academic origin (for which the earliest date of introduction was 1964 and the average date was 1979), the earliest of the commercial programs was developed in 1981/82, a date that clearly coincides

<sup>9</sup> In principle, in the aftermath of the (1981) *Diamond v. Diehr* decision, patent protection might also be available for some features of econometric software. In this area, as in many other software areas, there is tremendous resistance to this idea on the part of existing players, perhaps because they are well aware of the nightmare that might ensue if patent offices were unacquainted with prior art in econometrics (as is, no doubt, currently the case).

<sup>10</sup> Unfortunately, it is not possible to identify precisely the nature of the seed money support for many of the packages from the histories supplied in Renfro (2003), other than the simple fact that the development took place at a university.

Table 1  
Econometric software packages

Type of seed funding	Total number of products	Number commercialized	Average lag to commercialization	Average date of introduction
Research grants or own research	20	16	9.4	1979
Quasi-governmental organization	5	4	16.4	1974
Private (for profit)	5	5	0.8	1984
Total or average	30	25	9.0	1979

with the introduction of the non-hobbyist personal computer. Notwithstanding the academic research origin of most of these packages, today no less than 25 out of the 30 have been commercialized, with an average commercialization lag of 9 years.

Reading the histories of these packages supplied in Renfro (2003), it becomes clear that, although many of them had more than one contributor, normally there was a ‘lead user’ who coordinated development, the identity of the ‘lead user’ occasionally changing as time passed. Most of the packages had their origins in the solution of a specific research problem (e.g. the development of LIMDEP for estimation of the Nerlove and Press logit model, or the implementation of Hendry’s model development methodology in PCGive), but were developed, often through the efforts of others besides the initial inventor, into more general tools.

These facts clearly reflect the development both of computing technology and of the market for these kinds of packages. As predicted by our model, growth in the market due to the availability of personal computers and the growth of the economics profession as whole has caused the early, largely open source development model of the 1960s to become privatized. Nevertheless, there remain five programs that are supplied for free over the Internet; of these three had their origins prior to 1980 and the other two are very recent. As our model in Section 3 suggests, not all of the individuals in the community shift to the private system, and the share of PD activities can well be between zero and 1. Interestingly, only one of the five is explicitly provided with a GPL attached. A quote from one of the author’s websites summarizes the motivation of those who make these programs available quite well:

Why is EasyReg free?

EasyReg was originally designed to promote my own research. I came to realize that getting my research published in econometric journals is not enough to get it used. But writing a program that only does the Bierens’ stuff would not reach the new generation of

economists and econometricians. Therefore, the program should contain more than only my econometric techniques.

When I taught econometrics at Southern Methodist University in Dallas in the period 1991–1996, I needed software that my graduate students could use for their exercises. The existing commercial software was not advanced enough, or too expensive, or both. Therefore, I added the econometric techniques that I taught in class first to SimplReg, and later on to EasyReg after I had bought Visual Basic 3.

Meanwhile, working on EasyReg became a hobby: my favorite pastime during rainy weekends.

When I moved to Penn State University, and made EasyReg downloadable from the web, people from all over the world, from developing countries in Asia and Africa as well as from Western Europe and the USA, wrote me e-mails with econometric questions, suggestions for additions, or just saying “thank you”. It appears that a lot of students and researchers have no access, or cannot afford access, to commercial econometrics software. By making EasyReg commercial I would therefore let these people down.

There are also less altruistic reasons for keeping EasyReg free:

- By keeping EasyReg free my own econometric work incorporated in EasyReg will get the widest distribution.
- I will never be able to make enough money with a commercial version of EasyReg to be compensated for the time I have invested in it.
- Going commercial would leave me no time for my own research.<sup>11</sup>

<sup>11</sup> This quotation is from Hermann Bierens’ website at <http://econ.la.psu.edu/~hbierens/EASYREG.HTM>.

Indeed, the second statement suggests that one reason to leave the software in the public domain was that the researcher's commercial profits were not large enough. Likewise, the third statement suggests that the researcher cared about research and this was an important reason for not privatizing it. This is suggestive of the fact that the individual displayed a relatively low utility of commercial profits vis-à-vis his preference for research, which in turn affected his choice of staying public. In sum, the model's prediction that both private and public modes of provision can co-exist when at least some individuals adhere to community norms is borne out, at least for this one example.

We also discussed explicitly the role of complementary services or enhanced features for non-inventor users in the provision of software. This is clearly one of the motivations behind commercialization, as was illustrated by the example of *TeX*. Table 2, which is drawn from data in Renfro (2004), attempts to give an impression of the differences between commercialized and non-commercialized software, admittedly using a rather small sample. To the extent that ease of use can be characterized by the full WIMP interface, there is no difference in the average performance of the two types of software. The main differences seems to be that the commercialized packages are larger and allow both more varied and more complex methods of interaction. Note especially the provision of a macro facility to run previously prepared programs, which occurs in 84% of the commercial programs, but only in two out of the five free programs. Such programs are likely to require more user support and documentation, because of their complexity, which increases the cost of remaining in the PD system. In short, as our earlier discussion suggested, a commercial operation, which is likely to imply higher

profits, also provides a greater degree of additional investments beyond the mere availability of the research inputs.

To summarize, the basic predictions of our model, which are that participants in an open science community will defect to the private (IP-using) sector when profit opportunities arise (e.g. when the final demand for the product grows, or IP protection becomes available), are confirmed by this example. We also find some support for the fact that commercial operations are likely to undertake more complementary investments than pure open source operations. We do not find widespread use of the GPL idea in this particular niche market yet, although use of such a license could evolve. In the broader academic market, Maurer (2002) reports that a great variety of open source software licenses are in use, both viral (GPL, LPL) and non-viral (BSD, Apache-CMU).

Finally, our model in Section 3 does not explicitly incorporate all the factors that are clearly important in the case of software and databases. Specifically, one area seems worthy of further development. We did not model the competitive behaviour of the downstream firms in the database and software industries. In practice, in some cases, there is competition to supply these goods, and in others, it is more common for the good to be supplied at prices set by a partially price-discriminating monopolist. We report the evidence on price discrimination for our sample briefly here.

Table 3 presents some very limited data for our sample of 30 econometric software packages. Of the 30, five are distributed freely and a further eight are distributed as services, possibly bundled with consulting (such sales are essentially all commercial); this is the 'added value' business model discussed earlier. Of the remaining 17, we were able to collect data from their websites for 15. Of these, only two did not price discriminate, three discriminate by the size and complexity of the problem that can be estimated, and 10 by the type of customer (aca-

Table 2  
Comparing non-commercial and commercial software

Features	Share of non-commercial (%)	Share of commercial (%)
Full windows, icons, menus interface (WIMP)	60	60
Interactive use possible	60	68
Macro files can be executed	40	84
Manipulate objects with icons/menus	60	88
Generate interactive commands with icons/menus	20	60

Table 3  
Price discrimination in econometric software

Price discriminate?	No. of packages
By size or complexity	3
Academic/commercial	10
No discrimination	2
NA	2
Sold as a service	8
Free	5
Total	30



ademic or commercial).<sup>12</sup> A number of these packages were also offered in ‘student’ versions at substantially lower prices, segmenting the market even further. This evidence tends to confirm that in some cases, successful price discrimination is feasible and can be used to serve the academic market while covering some of the fixed costs via the commercial market.

Although price discrimination is widely used in these markets, it does have some drawbacks as a solution to the problem of software provision. The most important one is that features important to academics, or even programs important to academics, may fail to be provided or maintained in areas where there is either a very small commercial market or no market, because their willingness to pay for them is much lower. Obviously this is not a consequence of price discrimination per se, but simply of low willingness to pay; the solution is not to eliminate price discrimination, but to recognize that PD production of some of these goods is inevitable. For example, a database of elementary particle data has been maintained by an international consortium of particle physicists for many years. Clearly such a database has little commercial market.

## 7. Conclusions

Among the activities that constitute academic research, the production of software and databases for research purposes is likely to be especially subject to under-provision and privatization. The reason is, that like most research activities, the public good nature of the output leads to free riding, but that the usual norms and rewards of the ‘Republic of Science’ are less available to their producers and maintainers, especially the latter. In this paper we presented a model that illustrates and formalizes these ideas and we used the model to show that the GPL can be a way to ensure the provision of some of these goods, at least when the potential producers also want to consume them.

Although in this paper we have emphasized the beneficial role of the GPL as a coordination device for producing the public good, in these conclusions we also want to point out that the GPL is not a panacea that works in all situations, and one of those situations may indeed be the

production of scientific software and databases. One reason is that in practice it is difficult to distinguish between the ‘upstream’ activities, which, as we discussed, ought to be produced under public domain, and the ‘downstream’ ones. As we noted in the paper, the latter may entail important complementary investments. Therefore, they could be more effectively conducted under private rules that enable the producers to raise the rents that are necessary to perform such investments. However, the GPL ‘forces’ the contributors to work under public domain rules. If one cannot properly distinguish between upstream and downstream activities, the downstream activities, with implied complementary investments, will also be subject to public domain rules. This makes it more difficult to raise the resources needed to make the investments, with implied lower quality of the product.

To return to the example of the introduction, the *TeX* Users’ Group reports the following on their website in answer to the FAQ, “If *TeX* is so good, how come it’s free?”:

Its free because Knuth chose to make it so. He is nevertheless apparently happy that others should earn money by selling *TeX*-based services and products. While several valuable *TeX*-related tools and packages are offered subject to restrictions imposed by the GNU general public license (‘Copyleft’), *TeX* itself is not subject to Copyleft (<http://www.tug.org>).

Thus, part of the reason for the spread of *TeX* and its use by a larger number of researchers than just those who are especially computer-oriented is the fact that the lead user chose not to use the GPL to enforce the public domain, enabling commercial suppliers of *TeX* to offer easy-to-use versions and customer support.

The so-called ‘lesser’ GPL (LGPL) or other similar solutions can in part solve the problem. As discussed by Lerner and Tirole (2002), among others, the LGPL and analogous arrangements make the public domain requirement less stringent. They allow for the mixing of public and private codes or modules of the program. As a result, the outcome of the process is more likely to depend on the private incentives to make things private or public, and this might encourage the acquisition of rents in the downstream activities. Yet following the logic of our model, as we allow for some degree of privatization, the efficacy of the license as a coordination mechanism is likely to diminish. We defer a more thorough assessment of this trade-off to future research. Here, however, we want to note that when the importance of complementary investments is higher, one would expect LGPL to be socially more desirable. The benefits of having the downstream investments may offset the disadvan-

<sup>12</sup> The average ratio of commercial to academic price was 1.7. Assuming an iso-elastic demand curve with elasticity  $\eta$  and letting  $s$  = share of commercial (high demand) customers, one can perform some very rough computations using the relationship  $\Delta Q/Q = -\eta \Delta P/P$  or  $(1 - s) = \eta 0.7$ . If  $\eta = 1$ , then the implied share of academic customers is 70%. If the share of academic customers is only 30%, then the implied demand elasticity is about 0.42.

tage of a reduced coordination in the production of the public good. By contrast, when such investments are less important, or the separation between upstream and downstream activities can be made more clearly (and hence one can focus the GPL only on the former), a full GPL system is likely to be socially better.

**Acknowledgements**

Conversations with Paul David on this topic have helped greatly in clarifying the issues and problems. Both authors acknowledge his contribution with gratitude; any remaining errors and inconsistencies are entirely our responsibility. We are also grateful to Jennifer Kuan for bringing some of the open source literature to our attention.

**Appendix A. Model of public domain versus proprietary research**

*A.1. Set-up and equilibrium*

The total (indirect) utility of a researcher is  $U = z + \theta X(n - 1)$ , where  $X(n - 1)$  is the stock of PD knowledge when  $n - 1$  other researchers work under PD, and  $\theta \geq 0$  is a parameter that measures how much they care about the fact that others work under PD. Also,  $z = x(n)$  if they work under PD, and  $z = \pi$  if they work under PR, where  $x(n)$  is the utility that the researcher gains from her public contribution (assumed to be a function of the number of PD researchers  $n$ ) and  $\pi$  is the utility from the monetary income. We assume that  $x(n) \geq 0$ , and we make no assumption about the impact of  $n$  on  $x$ . There could be diminishing returns, that is a larger  $n$  implies smaller utility from one’s own contribution (e.g. because fewer important discoveries can be made), or there could be externalities, viz.  $x$  increases with  $n$ , or both. Note that we assume that the researchers enjoy the public contribu-

tion of the others even if they work under PR. We could make more complicated assumptions, e.g. the impact of  $X(n - 1)$  on utility is different according to whether the individual operates under PD or PR, but this will not affect our main results.

A researcher will produce under PD if  $\pi \leq x(n)$ . We assume that the individuals are heterogeneous in their preferences of PD versus PR, viz.  $[\pi - x(n)] \sim F(\cdot|n)$ , where the distribution function  $F$  depends on  $n$  because of  $x(n)$ . In principle, the support of  $\pi - x(n)$  is the whole real line. The share of individuals working under PD is then  $F(0|n)$ , and the equilibrium number of researcher  $n^e$  working under PD is defined by  $F(0|n^e) = n^e/N$ , where  $N$  is the total number of researchers in the community. This condition says that in equilibrium the share of researchers working under PD is equal to the share of researchers whose utility from  $\pi$  is not larger than the utility of their contribution  $x(n^e)$  to PD.

Fig. 1 depicts our equilibria. Point  $E$  in Fig. 1 is an equilibrium because if the number of researchers working under PD increases beyond  $n^e$ , the share of researchers with  $\pi - x(n) \leq 0$  increases by less than the share of researchers working under PD. But this is a contradiction because for some of the researchers who have moved to PD it was not profitable to do so. The reasoning is symmetric for the deviations from PD to PR in equilibrium. Stability of the equilibrium requires that the  $F(n^e)$  curve cuts the  $n^e/N$  line from above. This ensures that whenever an individual deviates from the equilibrium, moving from PD to PR, the share of individuals with  $\pi \leq x(n^e - 1)$ , i.e. those who find it profitable to operate under PD, is higher than the actual share of individuals working under PD after the move, viz.  $(n^e - 1)/N$ . Hence, the move is not profitable. Similarly, whenever an individual moves from PR to PD in equilibrium, the share of researchers with  $\pi \leq x(n^e + 1)$  becomes smaller than the share of researchers who now work under PD, viz.  $(n^e + 1)/N$ .

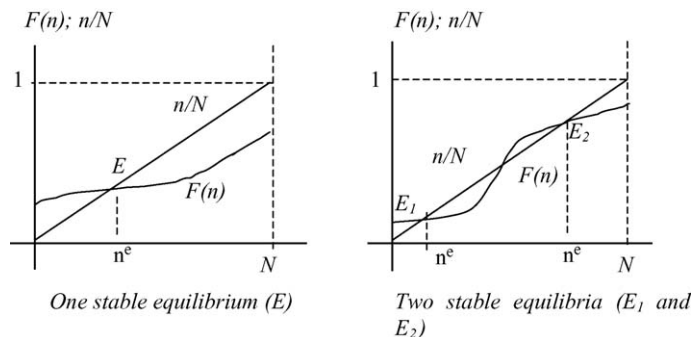


Fig. 1. Equilibria.

The stability conditions are then  $F(0|n^e - 1) > (n^e - 1)/N$  and  $F(0|n^e + 1) < (n^e + 1)/N$ . Multiple equilibria are also possible. There may be more than one  $n^e$  that satisfies (1) with  $F(n)$  cutting  $n/N$  from above, as shown by Fig. 1.

The share of researchers working under PD decreases if the economic profitability of research increases relatively to the researchers' utility from their public contributions. This can be thought of as a first order stochastic downward shift in  $F(\cdot)$  which would stem from an increase in  $\pi - x(n)$  for all the individuals. Likewise, a stronger taste for research would be represented by an upward shift in  $F$  as  $\pi - x(n)$  decreases for all the individuals. This raises  $n^e$ .

### A.2. Instability of PD knowledge production

To see why the production of knowledge under PD is unstable, suppose that  $n^e$  is an equilibrium and  $v$  researchers working under PR coordinate to work under PD. If  $n^e$  is an equilibrium, then  $\pi > x(n^e + v)$ , for at least one of these researchers, otherwise  $n^e + v$  would be an equilibrium. Yet, it is possible that  $x(n^e + v) + \theta X(n^e + v - 1) > \pi + \theta X(n^e - 1)$  for all the  $v$  researchers; viz., if the  $v$  researchers coordinate, they are better off than in equilibrium. To see this, recall that  $x(n) \equiv X(n) - X(n - 1)$ . Therefore,  $x(n^e + v) + \theta X(n^e + v - 1) - \theta X(n^e - 1) = x(n^e + v) + \theta \sum_{j=1}^v x(n^e + v - j)$ . But the expression in the summation sign is non-negative because  $x \geq 0$ . Hence, this expression can be greater than  $\pi$  in spite of the fact that  $\pi > x(n^e + v)$ . If the  $v$  researchers working under PR coordinate to work under PD, the system is unstable because at least one of them can find it profitable to deviate since he exhibits  $\pi > x(n^e + v)$ . Once he deviates, at least one of the remaining  $v - 1$  researchers has an incentive to deviate because  $n^e + v - 1$  is not an equilibrium, and so on till all the  $v$  researchers have deviated. At that point nobody else has an incentive to deviate because  $n^e$  is an equilibrium.

### A.3. The GPL model

Let  $B$  be an opportunity cost faced by an originator and any potential contributor to a project, with  $[B - x(n)] \sim G(\cdot|n)$ . We assume that even if the contributors do not join the project, they still enjoy  $\theta X(n)$  from the project if  $n$  researchers work on it under PD; that is, their indirect utility is  $B + \theta X(n)$ . Let  $n^G$  and  $n^{NG}$  be the equilibrium number of contributors joining the project under PD if the originator launches the project, works

under PD and attaches a GPL or not. Finally, let  $\tilde{n}^{NG}$  be the number of contributors under PD if the originator launches the project under PR (and cannot attach a GPL to it).

If the originator launches the project, works under PD and does not issue a GPL, the contributors working under PD will exhibit  $\pi \leq x(n^{NG} + 1)$  and  $B \leq x(n^{NG} + 1)$ . If  $\Gamma(\cdot)$  is the joint distribution function of  $\pi - x$  and  $B - x$ , in equilibrium we have  $\Gamma(0, 0|n^{NG}) = n^{NG}/N$ . With no GPL, the condition becomes less restrictive, because only  $B \leq x$  is required. As a result, with a GPL,  $G(0|n^{NG}) > n^{NG}/N$ . This will induce some researchers to join under PD, because the share of researchers with  $B \leq x$  is smaller than the share of researchers actually working under PD, i.e.  $n^{NG}/N$ . Provided that the stability conditions discussed above hold, the movement towards PD will stop at  $n^G$  such that  $G(0|n^G) = n^G/N$ . This implies  $n^G \geq n^{NG}$ , that is a GPL induces more researchers to work under PD.

If the originator launches the project under PR, the contributors can still join under PD. These will be all those with  $\pi - x(\tilde{n}^{NG}) \leq 0$  and  $B - x(\tilde{n}^{NG}) \leq 0$ . The difference with the previous no-GPL case is only that the originator does not join the project under PD. More generally, the same reasoning as above applies here, and  $n^G \geq \tilde{n}^{NG}$ . As a matter of fact, if  $n$  is large enough,  $\tilde{n}^{NG} \approx n^{NG}$ .

Given the behaviour of the contributors, will the originator who launches the project working under PD issue a GPL? With PD-GPL his utility will be  $x(n^G + 1) + \theta X(n^G)$ . With PD and no GPL it will be  $x(n^{NG} + 1) + \theta X(n^{NG})$ . By using the fact that  $x(n) \equiv X(n) - X(n - 1)$ , the former will be higher than the latter if  $X(n^G + 1) - (1 - \theta)X(n^G) \geq X(n^{NG} + 1) - (1 - \theta)X(n^{NG})$ . A sufficient condition for this inequality to hold is that  $\theta \geq 1$ . This follows from  $n^G \geq n^{NG}$  and the fact that  $X(n)$  increases with  $n$ , which in turn follows from  $x(n) \geq 0$ . Thus, if the originator chooses to work under PD, setting a GPL will be a dominant strategy unless  $\theta$  is close to zero (i.e. the impact of the others' behaviour is not that important) and some special conditions occur. For simplicity, we assume that  $\theta$  is large enough, and therefore choosing a GPL always dominates when the originator chooses PD.

If the originator chooses PR, his utility will be  $\pi + \theta X(\tilde{n}^{NG})$ . As a result, the originator will choose to work on the project under PD (and issue a GPL) if  $x(n^G + 1) + \theta X(n^G) \geq B$ , and  $x(n^G + 1) + \theta X(n^G) \geq \pi + \theta X(\tilde{n}^{NG})$ . If there were no GPL, the condition would be the same with  $n^{NG}$  in lieu of  $n^G$ . Since  $n^G \geq n^{NG}$ , with no GPL the condition becomes more restrictive. As a result, the possibility to use a GPL implies not

only that more researchers will join under PD, but also that more projects will be launched under PD with a GPL.

As discussed in the text, the GPL is least effective when there is a strong positive correlation between  $B$  and  $\pi$ . This implies that many individuals with small  $B$  tend to have a small  $\pi$  as well. As a result, the restriction  $\pi \leq x$  associated with  $B \leq x$  does not restrict the set of PD researchers much more than  $B \leq x$  alone, which means that  $n^G$  is close to  $n^{NG}$ , and the additional set of PD researchers created by the GPL is not large. In turn, this implies that the GPL does not encourage a more intensive coordination than without it.

The GPL raises the number of contributors working under PD in spite of the fact that the total number of contributors to the project decreases. To see this, assume for simplicity that  $x$  is roughly constant with respect to  $n$ , so that  $x(n^{NG}) \approx x(n^G) \equiv x$ . Consider the following table:

Researchers set	Action under no GPL	Action under GPL
$B \leq x; \pi \leq x$	Join under PD	Join under PD
$B \leq x; \pi \geq x (\pi \geq B)$	Join under PR	Join under PD
$B \geq x; \pi \leq x (\pi \leq B)$	Not join	Not join
$B \geq x; \pi \geq x$	Join under PR if $\pi \geq B$ Not join if $\pi \leq B$	Not join

Thus, with a GPL, some researchers who joined under PR switch to PD, while the opposite is not true. The researchers who no longer join the project with the GPL are only those who joined under PR. Thus, they do not affect  $n$  in equilibrium.

## References

- Allen, R.C., 1983. Collective invention. *Journal of Economic Behavior and Organization* 4, 1–24.
- Anton, J.J., Yao, D.A., 2002. The sale of ideas: strategic disclosure, property rights, and contracting. *Review of Economic Studies* 67, 585–607.
- Arrow, K., 1962. Economic welfare and the allocation of resources for invention. In: Nelson, R.R. (Ed.), *The Rate and Direction of Inventive Activity*. Princeton University Press, Princeton, NJ, pp. 609–625.
- Cohen, W.M., Florida, R., Randazzese, L., Walsh, J., 1998. Industry and the academy: uneasy partners in the cause of technological advance. In: Noll, R. (Ed.), *The Future of the Research University*. Brookings Institution Press, Washington, DC.
- Collins, S., Wakoh, H., 1999. *Universities and Technology Transfer in Japan: Recent Reforms in Historical Perspective*. University of Washington and Kanagawa Industrial Technology Research Institute, Japan.
- Dalle, J.M., 2003. Open source technology transfer. In: *Proceedings of the 3rd EPIP Conference, 22–23 November, Maastricht, The Netherlands*.
- Dasgupta, P., David, P.A., 1994. Toward a new economics of science. *Research Policy* 23, 487–521.
- David, P., 1993. Knowledge, property, and the system dynamics of technological change. In: *Proceedings of the World Bank Annual Conference on Development Economics 1992*, The World Bank, Washington, DC.
- David, P.A., 2002. The economic logic of open science and the balance between private property rights and the public domain in scientific data and information: a primer. In: *National Research Council Symposium on The Role of the Public Domain in Scientific and Technical Data and Information*, National Academy of Sciences, Washington, DC.
- David, P.A., 2000. A tragedy of the public knowledge ‘commons’? Global science, intellectual property and the digital technology boomerang. *Stanford Institute for Economic Policy Research Discussion Paper No. 00-02 (October) 2000*. Oxford Intellectual Property Research Centre-Electronic Journal/Working Paper Series. <http://www.oiprc.ox.ac.uk/EJWP0400.pdf>.
- European Commission, 2002. Draft directive on the patentability of computer-implemented inventions. Available at [http://www.europa.eu.int/comm/internal\\_market/en/indprop/comp/index.htm](http://www.europa.eu.int/comm/internal_market/en/indprop/comp/index.htm).
- Foray, D., Hilaire-Perez, L., 2005. The economics of open technology: collective organization and individual claims in the ‘Fabrique Lyonnaise’ during the old regime. In: Antonelli, C., Foray, D., Hall, B.H., Steinmueller, W.E. (Eds.), *Essays in Honor of Paul A. David*. Edward Elgar.
- Geuna, A., Nesta, L., this issue. University patenting and its effects on academic research: the emerging European evidence. *Research Policy*, doi:10.1016/j.respol.2006.04.005.
- Hall, B.H., 2004. On copyright and patent protection for software and databases: a tale of two worlds. In: Granstrand, O. (Ed.), *Economics, Law, and Intellectual Property*. Kluwer Academic Publishers, Boston.
- Hall, B.H., Link, A.N., Scott, J.T., 2001. Barriers inhibiting industry from partnering with universities. *Journal of Technology Transfer* 26, 87–98.
- Hall, B.H., Ziedonis, R., 2001. The determinants of patenting in the U.S. semiconductor industry, 1980–1994. *Rand Journal of Economics* 32, 101–128.
- Harhoff, D., Henkel, J., von Hippel, E., 2003. Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations. *Research Policy* 32, 1753–1769.
- Heller, M.A., Eisenberg, R.S., 1998. Can patents deter innovation? The anticommons in biomedical research. *Science* 280, 698–701.
- Hertzfeld, H.R., Link, A.N., Vonortas, N.S., this issue. Intellectual property protection mechanisms and research partnerships. *Research Policy*.
- Isabelle, M., 2004. They invent (not patent) like they breathe: what are their incentives to do so? Short tales and lessons from researchers in a public research organization. In: *Proceedings of the Third EPIP Workshop, 2–3 April, Pisa, Italy*.
- Kuan, J., 2002. *Open Source Software as Lead User’s Make or Buy Decision: a Study of Open and Closed Source Quality*. Stanford University.
- Lee, Y.S., 2000. The sustainability of university–industry research collaboration. *Journal of Technology Transfer* 25, 111–133.
- Lerner, J., Tirole, J., 2002. Some simple economics of open source. *Journal of Industrial Economics* 50, 197–234.
- Lerner, J., Tirole, J., 2005. The scope of open source licensing. *Journal of Law Economics and Organization* 21, 20–56.
- Levinthal, D., March, J.G., 1993. The myopia of learning. *Strategic Management Journal* 14, 95–112.

- Maurer, S.M., 2002. Promoting and disseminating knowledge: the public/private interface. Working paper.
- Merton, R.K., 1957. Priorities in scientific discovery: a chapter in the sociology of science. *American Sociological Review* 22, 635–659.
- Merton, R.K., 1968. The Matthew effect in science. *Science* 159 (3810), 56–63.
- Nelson, R.R., 1959. The simple economics of basic scientific research. *Journal of Political Economy* 77, 297–306.
- Nuvolari, A., 2004. Collective invention during the British Industrial Revolution: the case of Cornish pumping engine. *Cambridge Journal of Economics* 28, 347–363.
- Olson, M., 1971. *Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, Cambridge, Massachusetts.
- Owen-Smith, J., Powell, W.W., 2001. To patent or not: faculty decisions and institutional success at technology transfer. *Journal of Technology Transfer* 26, 99–114.
- Renfro, C.G., 2003. Econometric software: the first fifty years in perspective. *Journal of Economics and Social Measurement* 29, 1–51.
- Renfro, C.G., 2004. A compendium of existing econometric software packages. *Journal of Economics and Social Measurement* 29, 359–409.
- Scotchmer, S., 1991. Standing on the shoulders of giants: cumulative research and the patent law. *Journal of Economic Perspectives* 5, 29–41.
- Stephan, P.E., 1996. The economics of science. *Journal of Economic Literature* 34, 1199–1235.
- TeX Users' Group, 2005. Website <http://www.tug.org>.
- Thursby, J., Jensen, R., Thursby, M.C., 2001. Objectives, characteristics and outcomes of university licensing: a survey of major U.S. universities. *Journal of Technology Transfer* 26, 59–72.
- Von Hippel, E., 1987. Cooperation between rivals: informal know-how trading. *Research Policy* 16, 291–302.
- Von Hippel, E., 1988. *The Sources of Innovation*. Oxford University Press, Oxford.
- Von Hippel, E., von Krogh, G., 2003. Open source software and the private-collective innovation model: issues for organization science. *Organization Science* 14, 209–223.