Available online at www.sciencedirect.com
SCIENCE DIRECT®
Research Policy 35 (2006) 875–892
research policy
www.elsevier.com/locate/respol

Proprietary versus public domain licensing of software and research products

Alfonso Gambardella [a], Bronwyn H. Hall [b,c,*]

[a] *IEGI, Bocconi University, Via Filippetti 9, Milan 20122, Italy*
[b] *Department of Economics, University of California at Berkeley, Berkeley, CA 94720-3880, USA*
[c] *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.

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 (B.H. Hall).

0048-7333/$ – see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.respol.2006.04.004

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*.[1] 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.[2] 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

[1] 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).

[2] WYSIWYG is a widely used acronym in computer programming design that stands for "what you see is what you get".