

UC Irvine

UC Irvine Previously Published Works

Title

Intellectual Property and eScience

Permalink

<https://escholarship.org/uc/item/4t3190sm>

Author

Burk, DL

Publication Date

2007

Peer reviewed

Intellectual Property in the Context of e-Science

Dan L. Burk

Law School
University of Minnesota

E-science promises to allow globally-distributed collaboration and access to scientific research via computer networks, but e-science development is already encountering difficulty over the intellectual property rights associated with data and networked collaborative activity. The proprietary nature of intellectual property is generally problematic in the practice of science, but such difficulties are likely to be exacerbated in the context of e-science collaboration where the development and use of intellectual resources will likely be distributed among many researchers in a variety of physical locations, often spanning national boundaries. While a potential solution to such problems may reside in the mechanism of "open source" licenses, the organizational structure of scientific research may not map cleanly onto the open source model. Consequently, a firm understanding of not only the technical structure but of the social and communicative structure of e-science will be necessary in order to adapt licensing solutions to the practice of e-science.

doi:10.1111/j.1083-6101.2007.00340.x

Introduction

The advent and proliferation of global computer networks have altered the practice of science, and additional changes seem sure to come. Scientists already routinely collaborate and access informational resources by way of the Internet and associated technologies (Glasner, 1996). Further advances in this direction are contemplated via so-called "grid" technologies to enable collaborative sharing of both information and resources on an international scale (Hey & Trfethen, 2003; Newman, Ellisman, & Orcutt, 2003). Such distributed computing architectures promise to make available processing power, data storage, and related large-scale computing resources independent of geographic location. Researchers participating in such technological collaborations are increasingly drawn into distributed communities and far-flung alliances that might previously have been impossible (Finholt, 2002).

With the promise of such capabilities comes a variety of new challenges and, in particular, new social and cultural challenges to the conduct of research. A variety of

non-technical legal and social factors are likely to shape the structure, character, and success of such initiatives (Burk, 2000; David & Spence, 2003). Crucial aspects of these determinants are centered upon ownership and control of the information and tools associated with e-science. While it is by no means the only legal regime that will affect these aspects of e-science, intellectual property law is expected to play a major role in determining such ownership and control (Eckersley, Egan, & Shun-ichi, 2003).

For example, intellectual property issues were recently identified as a serious concern in the organization and conduct of the eDiaMoND (Digital Mammography National Database) project in the United Kingdom (Hinds, Jirotko, Rahman, D'Agostino Meyer Piper et al. 2005D'Agostino, Meyer, Piper, et al., 2005). This project employs collaborative computing or "grid" technologies to allow collection and sharing of digitized mammograms and associated patient data acquired from a nationwide screening program. Anonymized medical radiographic images from the project were digitized and transferred to a computer database accessible for data mining or other algorithmic analysis. Yet ownership and control of the individual images and the collective database remained indeterminate; possible claimants to ownership or control of the images might include the patients, their physicians, radiographic technicians, sponsoring institutions, or even the database engineers.

In previous work, I identified a range of intellectual property issues, primarily arising out of Internet-based legal disputes that seemed likely to become issues for the then-nascent e-science phenomenon (Burk, 2000). Since that time, practices associated with e-science have advanced dramatically, although often along paths that were not entirely anticipated. Early predictions regarding formalized "collaboratory" research environments have yet to be entirely realized, although experiments in such shared virtual spaces continue. At the same time, the availability of publicly shared research resources, using the readily available medium of the Internet, has grown. New technological architectures for interactive grid computing resources are under development, but perhaps most surprising has been the mix of cultural and legal influence of open source licensing models to facilitate open access to scientific resources.

In this article, I expand and update earlier work to take account of such developments. I begin by reviewing the basic forms of intellectual property most relevant to e-science and their likely application to the tools, methods, and products of e-science. I also discuss the shift in approach towards intellectual property observed in the scientific community over the past two decades. I then turn to the likeliest impediment posed by the confluence of these technological and social factors, which is the jurisdictional inconsistency faced by distributed research collaborations. I discuss contractual solutions to these impediments, especially the spread of open source licensing, not only as a formal legal mechanism to promote e-science but also as a normative statement about the culture and ethos of research practice. I conclude by considering the social and organizational challenges posed by open source licensing itself and the prospects for addressing the transition of this legal mechanism from its original milieu to that of e-science.

Intellectual Property

Legal ownership and control over the products of innovative and creative effort have typically been recognized as one or more forms of intellectual property rights. Several of these different forms of intellectual property have become directly relevant to the practice of e-science, as e-science may generate valuable ideas, methodologies, data, databases, texts, software, and even tangible research tools.

Much of the intellectual property related to e-science will fall under the copyright system. Copyright arises spontaneously upon fixation of an original work in a tangible medium of expression, and generally lasts for the lifetime of the author plus 70 years. The rights that attach at fixation include not only the exclusive right to reproduce the work but also some version of the exclusive rights to adapt, distribute, publicly perform, and publicly display the work. Copyright also includes the right to exclude others from reproducing, adapting, distributing, publicly performing, or displaying works that are substantially similar to the protected work. In most countries, these rights are subject to some exceptions or user privileges, such as rights to use short quotations, or to criticize or comment upon the work without permission of the copyright holder.

Restricting uses of the particular expression but allowing ideas to be freely re-used, copyright extends only to the original expression found in the work and not to the idea expressed. Copyright also expressly eschews protection of facts, processes, machines, and functional or utilitarian items. Traditionally, copyright has focused on aesthetic or artistic works, but these categories have come to include texts that “behave,” including software. More traditional texts, such as scientific reports and journals, are also covered by copyright law. Consequently, copyright has become increasingly important to technological innovation as well as to artistic creativity.

Copyright may also offer some degree of protection to databases since it protects the original and creative selection and arrangement of compilations. Frequently, individual data in a database will constitute unprotectable facts, but their selection and arrangement may be covered by copyright. Nonetheless, copyright often offers only relatively weak protection for databases, especially if the selection and arrangement of the data are dictated by the use or structure of the data itself or otherwise dictated by technical or organizational constraints. In such cases, which are frequent, the selection and arrangement may not be considered “original” and so may fail to meet the criteria for copyright protection.

Patents are a second form of intellectual property that has become increasingly relevant to e-science. Patent law has traditionally been the form of intellectual property directed to functional or utilitarian items: processes, machines, articles of manufacture, and compositions of matter. Patent law excludes from protection products of nature, that is, materials that have not been in some way changed or altered by humans. Patent law also typically excludes from its ambit laws or principles of nature, including pure mathematics, on the theory that these are discoveries inherent in the natural world rather than the products of human ingenuity.

Unlike copyright and trade secrecy, which arise spontaneously based on the nature and use of the protected information, patents entail rights that arise only after the review and approval of a governmental agency. Inventors must apply for a patent, submitting for examination a document demonstrating that the invention meets the statutory criteria of novelty, usefulness, and inventiveness. The application must also disclose how to make and use the invention and must include claims that set forth the characteristics of the technology to be covered by the patent. The examining agency will reject applications that fail to meet these criteria, requiring applicants to either amend their claims to meet the requirements or forgo issuance of a patent.

Because the patent specifies in its claims the limits of the technology it covers, the coverage of a patent may be broad or narrow, depending upon the scope of the claims. Unlike the fairly sharply delineated acts excluded by copyright, the range of infringing acts covered by patent rights is quite broad and expansive. Once approved, the patent confers the exclusive rights to make, use, sell, offer for sale, or import the invention described in the patent claims. Also unlike copyrights, patents typically are subject to very few user privileges or exceptions.

Although patent and copyright are generally directed to different subject matter, the functional and the expressive, they meet in certain technologies that may be considered expressive. While the symbolic indicia of computer code are protected under copyright, functional aspects of software may be subject to patent protection. Most major jurisdictions now recognize software as patentable in some form. In the United States, this rationale has been taken quite far, first to apply patents to any task or method programmed as a computer process and more recently to apply patents to any process with a useful result. Other jurisdictions, such as those belonging to the European Patent Convention, are more restricted to methods or processes in their application of patents.

In addition to patents and copyrights, considerations of trade secrecy may be important to e-science collaboration. Trade secrecy entails any type of business information that is not generally known but which confers a business advantage. Trade secrecy frequently arises out of contractual obligations; the parties to a business relationship may specify in a contract that certain information is proprietary and will be kept confidential. Trade secrecy sometimes also arises out of other legal duties, such as a duty of loyalty to an employer, or a fiduciary duty to a partner in a business alliance, or from a legal prohibition on improperly profiting from another's efforts.

Finally, the European Union (E.U.) and some other jurisdictions recognize a form of intellectual property right in collections of information, or databases. Because copyright offers little protection for many types of databases, some nations have sought to foster their domestic database industries with novel forms of database protection. The E.U. has been the global leader in this trend by way of a database directive requiring the member states to enact a *sui generis*, or novel form of intellectual property protection, prohibiting the unauthorized extraction of information from proprietary collections of information. This directive includes a reciprocity provision, requiring that other nations offer a database protection in their jurisdiction

in order for their nationals to gain the benefit of the database right in the E.U. As a consequence, there is considerable incentive for trading partners of the E.U. member states to enact reciprocal legislation.

Open Science

Intellectual property rights are typically justified on the utilitarian theory that law should provide an economic incentive in order to foster creativity and innovation. Unlike tangible goods, the intangible subject matter of intellectual property is easily copied and appropriated. Consequently, it is difficult for artists and inventors to exclude others from freely appropriating their creations; this in turn means that it is difficult to charge payment for use of the creation. Intellectual property laws provide legal exclusion to mimic the physical exclusion that inheres naturally in tangible objects or real property. Creators can use legal process to exclude users or to allow access on condition of payment. The opportunity to exclude, and so to charge payment, is thought to provide a monetary incentive for more creation.

The scientific research enterprise, however, has traditionally been viewed as operating on quite a different incentive mechanism, that of open access and reputational reward. This system of open science was classically articulated by Robert Merton in the mid-20th century (Merton, 1973). By examining the community of scientific researchers *as* a community, Merton identified a variety of behavioral expectations that scientists claimed to be important to their work and to their interactions. The primary scientific norms that emerged from these studies include:

- *universalism*: the expectation that scientists should judge empirical claims according to impersonal criteria, without regard to the identity of their author;
- *disinterestedness*: the expectation that scientists will subordinate their own biases and interests to the advancement of knowledge;
- *communalism*: the expectation that discoveries will be freely shared and dedicated to the community of scientists; and
- *organized skepticism*: the expectation that scientists will subject empirical claims to systematic scrutiny and validation.

Later he added an additional norm of *originality*: the expectation that contributions to the fund of scientific knowledge would be valued for their novelty.

As identified by Merton, these norms function together to drive the scientific enterprise, creating community behavior that reputationally rewards contributions to the corpus of scientific information. Scientists are expected to contribute their discoveries freely to the community; such contributed knowledge is vetted through criticism and peer review of published papers or reports. Thus, publication of new scientific discoveries instantiates both the values of communalism and organized skepticism. Contributed knowledge that passes such scrutiny gains for the contributor the recognition and respect of his or her peers. Under the norm of universalism, any contributor to the fund of scientific knowledge can expect to receive such

recognition, regardless of social status; the value of disinterestedness ensures that all contributions will be treated on an equal footing.

Merton's analysis has been the subject of subsequent criticism for both alleged theoretical and pragmatic failures—in particular, for accepting at face value the assertions of interviewed scientists regarding community norms, rather than tracking scientists' actual practices. Merton's framework nonetheless continues to resonate both with analysts and with the scientific community itself. In particular, the Mertonian framework has provided a useful lens through which to view the collision of open science and proprietary intellectual property. Several analysts have documented and discussed the clash between these two systems of reward (Eisenberg, 1987; Merges, 1996). Financial pressures on the academic system housing the scientific community have led to pressure toward patenting, especially in the life sciences, by college and university administrators. Academic institutions frequently lack the expertise or resources to develop adequately or license patented inventions, and the majority of such inventions in any event lack significant commercial value, but the occasional commercial "blockbuster" university patent fosters a gold-rush mentality toward patents (Rai, 2000). This mentality also attends the prospect for individual scientists to participate financially in significant royalties or to profit from the public offering of small companies built around patented technology, and it creates similar pressure on a personal level towards patenting.

Adherents to the more traditional set of Mertonian norms may view the financial reward system of intellectual property as distorting the research agenda of academic science and improperly commodifying information, techniques, and materials that should be freely accessible to the entire community (Nelson, 2004). On occasion, the community or its formal institutions may mobilize to enforce the traditional normative structure and frustrate the intrusion of financial incentives into the academic research enterprise. For example, in the name of open and communal science, many prominent scientific journals have demanded that as a condition of publication, researchers make openly available materials and data upon which their papers are based. Absent such a requirement for access, researchers might attempt to gain both the financial reward of the intellectual property system and the reputational reward of science, cheating the value of communality by keeping materials and background data proprietary through trade secrecy or patent licensing.

This clash of incentives has been apparent, for example, in the development of the international Human Genome Mapping project, where the sequence information generated under the project became the subject of both public and private attempts at patenting (Boyle, 2003). Early governmental efforts to patent nucleotide sequences generated from the project were vehemently rejected by the genetic sequencing community, primarily on the grounds that such patents would violate Mertonian norms of communality. Interestingly, the justification that patents could be used to keep the sequences publicly available by preempting private efforts to obtain patents was unpersuasive to critics of the governmental patenting effort. In the face of such criticism, the governmental patenting efforts were abandoned.

Subsequent patenting efforts by commercial entities were the subject of even greater criticism. Several private companies embarked upon ambitious genomic sequencing initiatives, keeping their sequencing data proprietary and filing patent applications on the sequences. In response, the publicly funded sequencing laboratories agreed on a collective patent-defeating strategy, requiring academic researchers to deposit their sequence data to publicly available repositories on the Internet within hours of the data generation. By making the data publicly accessible, the academic researchers intended not only to implement the norm of communality but to defeat the novelty of patent applications on the sequences. Agreement to these rules was largely the product of normative pressure by the research community in addition to some formal enforcement by scientific journals that required sequence disclosure as a stipulation for article publication.

However, the deposit of the publicly financed genomic data to open repositories also allowed private firms to enhance their proprietary databases by capturing the publicly available data, admixing it with their own data, and making the resulting database available on a commercial, restricted basis. The firms offering such data access ultimately abandoned such business strategies, in part due to the competition from the publicly available sequence repositories. Despite the failure of commercial genomic data provision, the private capture of the data in those instances alerted researchers to the danger of simply releasing communal products into the public domain; such a strategy may prevent patenting of the particular product but does not prevent its capture in some other closed, proprietary form (Burk, 2005).

Interjurisdictional Collaboration

The dispute over disposition of the information generated by the Human Genome Mapping project presages future disputes in e-science collaboration. In many aspects, intellectual property right disputes in e-science will be no different from those in traditional modes of scientific research; the culture clash over norms of openness versus proprietary ownership can occur in the isolated physical laboratory as well as on the distributed computing grid. Of course, new communicative and computational tools may increase the intellectual property emphasis in e-science. For example, copyright and database protection laws may loom somewhat larger in e-science, given the dependence of e-science on large data collections and on the software necessary to access and process such data sets.

Yet the legal landscape for e-science is likely to differ dramatically from traditional science in at least one aspect. In a global environment of computerized, distributed access and collaboration, application of territorially-based law becomes uncertain. Data located in one jurisdiction may be accessed nearly instantaneously and used in a different jurisdiction; the law governing such use may differ dramatically between such jurisdictions. Scientists located in different jurisdictions may jointly generate new data, techniques, or materials; the laws governing ownership and allocation of such collaborative products may again differ dramatically among

jurisdictions. Determining which jurisdiction's laws should govern is a difficult and often uncertain exercise.

Consider, for example, the intellectual property evaluation of the eDiaMoND project mentioned at the beginning of this article. The intellectual property concerns studied in relation to this U.K. project were addressed on a relatively parochial scale; data for the project were captured within a single jurisdiction, the United Kingdom, and accessibility was considered only within that jurisdiction. Preliminary legal and social analysis for the problem was conducted primarily under U.K. law, and recommendations for resolution of ownership and control were couched in terms of that jurisdiction's institutions. While recognizing the possibility that the data could be made available to researchers in other jurisdictions, the evaluation stayed relatively close to home.

Analysis of legal claims becomes far more complicated when data are generated and accessed across multiple jurisdictions, involving multiple actors and institutions (Burk, 1994). The relevant factors in such cases may include not only the laws of multiple nations but the policies and expectations of sponsoring institutions and practice communities. A first issue would be to decide where a legal claim regarding shared data might be adjudicated, that is, which nation's courts might have power to hear the case. Typically, a nation where the data are uploaded or downloaded, or whose nationals have some substantial involvement with the data, would be potential adjudicatory venues (Goldsmith, 1998). A court hearing the case would then have to decide which nation's laws should apply to the claim. Where the legal result in different jurisdictions would be inconsistent, the law includes a set of meta-rules, known as the law of conflicts, to decide which jurisdiction's laws will prevail. Of course, since jurisdictions may apply different legal rules to conflict analysis, meta-meta-rules may be applied to determine which jurisdiction's law of conflicts will prevail.

For example, intellectual property protection arises under national law, extending only as far as the border of a country that has granted a copyright, patent, or other form of protection. No international or multijurisdictional grant of intellectual property rights exists. Additionally, because the grant of intellectual property rights is unique to each country, the scope and character of such rights varies from country to country. International treaties have harmonized to some extent the characteristics of intellectual property rights around the world. The Berne Convention for the Protection of Literary and Artistic Works has, for well over a century, set minimum standards of copyright protection for signatory nations. Additionally, the treaty on Trade Related aspects of Intellectual Property (TRIPs) sets certain minimum standards for patents, copyrights, and other forms of intellectual property. Accession to TRIPs is a requirement for admission to the World Trade Organization (WTO), membership in which has been sought by the majority of nations (Burk, 1998).

However, significant idiosyncrasies of national law remain. Even within the TRIPs framework, the treaty sets the floor, but not the ceiling; signatory nations

must comply with certain minimum standards set out in the treaty, but there may be local jurisdictional variations away from this baseline. For example, in the E.U., which promulgated a directive intended to “harmonize” copyright law among the member states, users of copyrighted works have differing and sometimes inconsistent use privileges. Some nations are not signatories to TRIPs; as of this writing, some important jurisdictions, notably Russia, remain outside the WTO, and so outside the TRIPs framework. Least developed nations are also allowed some degree of variance from the requirements of TRIPs due to their economic status.

In the case of eDiaMoND, preliminary analyses of image ownership assumed that copyright inheres in the images, and the legal analysis proceeded from there. This is arguably the correct legal assumption to make under U.K. copyright law but is far more questionable under U.S. law; copyright is less likely in the U.S. to cover the images because the images constitute facts, which are excluded from copyright. U.S. law certainly recognizes copyright in the artistic arrangement or expression in photographic images, but it is unlikely that mammogram images are arranged so as to be artistic or expressive. Rather, the images were likely captured according to utilitarian criteria. To the extent that the arrangement of the images was dictated by functional considerations, American copyright law is unlikely to protect them. Thus, such images might well be unprotected under U.S. copyright law.

Additionally, the U.S. lacks the *sui generis* database right found in the U.K. and other E.U. nations, so that the database as a whole could only be protected in the U.S. by copyright. To the extent that the selection and arrangement of images in the database may not be original expression—that is, to the extent that it is dictated by functional or utilitarian considerations—it may fail U.S. copyright requirements. At a minimum, extracting data from the database in such a way so as not to take the selection or arrangement, for example extracting single images, might well avoid any copyright liability in the U.S.

Similarly, under U.S. patent law, a patent or printed publication disclosing an invention anywhere in the world will disprove the novelty of the invention, possibly precluding the inventor from obtaining a patent. However, the novelty of an invention can be disproved by “knowledge or use” of the invention only in the U.S., its territories, or possessions. Thus, whether online activity is classified as a “printed publication” or as “knowledge or use,” and whether the latter type of online activity is deemed to occur in the U.S., may affect the ability to patent an invention in the U.S. (Burk, 1993).

Patents for online collaboration may also be affected by other territorial provisions (Burk, 1993). U.S. law explicitly states that co-inventors need not work in the same physical vicinity in order to be named on a patent. However, U.S. law excludes from consideration for a patent inventive work that is not carried out either in U.S. territory or in a WTO country. Thus, inventive work that occurs in non-WTO jurisdictions, such as Russia, would not qualify for a U.S. patent. In an online collaboration between Russian and American researchers, determining the location of inventive work could be both critical and contested.

Another key patent law provision that is pertinent for open access research is the presence or absence of a research exemption to infringement. Many industrialized nations have enacted statutory experimental use exceptions to their patent law, allowing experimentation with and testing of patented inventions without permission of the patent owner. Canada, lacking such a statutory exception, has developed a strikingly broad common-law exception for experimental patent use. However, the U.S., outside of a very narrow statutory exception for development of regulatory health and safety data, has essentially no experimental use exception for commercial uses of a patented invention.

At the same time, in the copyright area, the U.S. has developed a strikingly broad and flexible fair use exception that is based on common-law and statutory provisions and is unique in the world. The majority of nations follow the statutory approach known in the U.K. and Canada as “fair dealing,” essentially a list of specific exemptions from copyright liability under specific circumstances. U.S. copyright law, rather than limiting exceptions to a particular list, assesses whether an unauthorized use is permissible or “fair” on a case-by-case basis, according to broad statutory criteria. Thus, the research use of a copyrighted work may be permissible in one jurisdiction while impermissible in another. This may be a particular problem for software, which may be covered by both patent and copyright law and for which the patent and copyright exemptions may be inconsistent both within and among jurisdictions.

Contractual Solutions

Interjurisdictional conflicts, although not an issue that research scientists have routinely confronted, have existed in some form for as long as there have been territorial borders and the means to transverse them. Businesses have long functioned under such conditions, in large part due to skill in contracting. Similarly, the differing expectations of e-science participants and the different legal regimes they will encounter can be addressed to some extent by contract. Parties to a contract can state the expected disposition of intellectual property and can often also specify the legal or jurisdictional regime under which the transaction will be treated.

However, contracts are necessarily incomplete. The parties to a contract are not prescient and cannot foresee every eventuality. Indeed, attempting to anticipate every possible contingency can become prohibitively expensive. Inevitably some situation will arise that has not been considered in the contract. In such situations where a dispute arises in the gaps of the contract, resolution of legal issues will revert to the default rules of property and jurisdiction, requiring either the parties or an adjudicatory body to navigate all the issues reviewed above.

Contracts are also limited in their application. Unlike property rights, which when recognized under a given legal regime are good against all comers, contracts are enforceable only against the parties to the contract. Additionally, under some circumstances, contracts may be legally unenforceable due to public policy restrictions on the right of contract. In order to protect their citizens or effectuate certain

policies, jurisdictions may forbid contracts with particular terms. For example, many European nations refuse to recognize contracts with choice of law provisions that would subject consumers to less protective law than their own. In the U.S., contract provisions that are enforced under state law may be preempted if they undermine or violate the policies behind federal copyright and patent statutes.

In the commercial sector, these limitations have been long recognized and have led to the proliferation of standardized form contracts. Such “off the shelf” forms streamline the costs of the contracting process and have replaced actual negotiation in every kind of consumer transaction, from automobile parking to cell phone service. Form contracts are offered on a “take it or leave it” basis, with non-negotiated “boilerplate” or “fine print” terms that consumers seldom read but may be required to honor if they manifest assent to be bound by the agreement.

Mass market licenses have become ubiquitous in, but only more recently have impinged upon, intellectual property, particularly as regards copyright in computer programs. In the context of software, such contracts appeared first as the so-called “shrink-wrap” license, which was purportedly triggered when the purchaser opened the packaging or shrink-wrap on a box containing software media. Since software is seldom distributed in such packaging any more, more recent incarnations of the license include the “click-wrap” license, in which the user triggers the license via mouse or keyboard by clicking on a graphic display that reads “I agree,” and the “browse-wrap” license, which purports to be triggered when the user accesses the website on which the license text is displayed.

Either as a substitute for or as an addition to the control of information afforded by intellectual property rights, standardized licensing has been feared as a potential threat to open science (Reichman & Uhler, 2003). Such licenses have been adopted in the context of laboratory research as standardized “letters licenses” for materials transfer in biotechnology research. Similar form contracts could be used to restrict the flow of information among online collaborators (Burk, 2000). Alternatively, creative uses of such licenses could perhaps secure the free flow of information in e-science, as has been done in the context of open source software licensing.

Open Source Licensing

One of the most debated and celebrated developments in collaborative work production has been the emergence of communal software projects facilitated by Internet communications and characterized by a philosophy of open collaboration and sharing. This “free and open source software” movement has developed as a response to commercial software production that strives to keep software proprietary via intellectual property or technical closure. Commercial firms often distribute only machine-readable versions of programs and keep the human-readable source code restricted. In contrast, the loose confederation of open source software producers holds that software source code should be freely available for others to critique, to tinker with, and to improve.

The role of normative constraints and reputational capital in open source software development suggests that it has much in common with the culture of science (Burk, 2005; de Laat, 2001). The norms of communalism, universalism, and organized skepticism all appear to be operative in the open source and free software community. Just as scientists publish their work to ensure that it remains openly accessible for others to build upon, so open source coders circulate their code, employing “copyleft” and similar licensing mechanisms to maintain accessibility. In open source, as in Mertonian science, peer review is relied upon to certify additions to the canon of knowledge. In both communities, normative expectations are largely communitarian rather than proprietary, and commercialization of work product is frequently viewed with suspicion. Both communities tend to rely on reputational incentives to prompt contributions to the common fund of knowledge. As scientists insist on freedom to set their research agendas, so open source programmers have insisted on freedom to work on the code that most stimulates their interest.

However, in a collaborative effort where open, normative constraints are the only impediment to deviant behavior, participants could selfishly defect from the community and incorporate the software into proprietary, closed products. Third parties might also obtain the openly available communally produced software and attempt to divert it into a closed form. Even if an open version remained available, the closed version could well become the dominant implementation of the code, crowding out the use and maintenance of the open version. This possibility could be particularly likely if the closed version were attached to or bundled with a dominant, proprietary platform, such as Microsoft Windows. Economic network effects and user “lock-in” accompanying a dominant product could serve to capture permanently code that began as open source.

This possibility threatens both the structure of the community and the development of its work product. Diversion of collaborative code would not only violate the philosophical tenets of the community but might also disrupt coding collaboration due to the threat of free-riding on work of the community; programmers would be less likely to contribute work to the project if they knew it may be exploited by someone else. Indeed, the threat to the open source project from large corporate software concerns, and most especially from Microsoft, is a dominant theme, and almost a phobia, in open source circles.

Yet over the past decade, the free and open source communities have, in a clever bit of legal jujitsu, adopted the mechanism of standardized form licensing for maintaining open access to their software projects. Open source coders have attached to their products a type of standardized form contract with a particular twist; as a term of copying, adapting, or re-distributing the software, the user must agree to maintain open access to the source code. These licenses come in a variety of configurations, with terms that may forbid commercialization of the code, or may allow commercialization, or may allow commercialization on certain specific terms. One particular form of license, known as copyleft licensing, requires any change or improvement of

the software to be made available to the public upon the same terms as received; in other words, the license has a “viral” character, attaching its terms to any products derived from the initially licensed software.

Thus the automatic license, in conjunction with copyright law, becomes an assurance of openness rather than of restriction. This by no means places open source projects in the public domain, at least not as that term is generally employed in discussing intellectual property. Typically, the public domain encompasses those categories of intangible goods not covered by intellectual property, but open source software is unquestionably the subject of intellectual property protection. Copyright protection for these projects has neither been waived nor disclaimed in any fashion; indeed, the currency of intellectual property protection for open source code is the premise upon which copyleft and other open source licenses are based. The threat behind the license is that a copyright suit will ensue if the code is used in any fashion other than the open access under the license.

At the same time, the mystique of open source has given rise to a mythology of “self-assembling” Internet projects, in which the invisible hand of cooperation coordinates worker output. Undoubtedly, the low-cost communications medium of the Internet does in fact allow open source coders to locate and coordinate projects more easily. Yet both anecdotal and empirical analyses suggest that the practice of open source coding is very tightly controlled by a few project leaders (McGowan, 2001; Tuomi, 2000). Strong normative expectations, backed by social sanctions, are directed at maintaining the focus and trajectory of open source projects. Project participants may vote with their feet; they may exit a project if they wish but typically cannot “fork” a project to develop code contrary to the direction set by leaders. If a programmer’s developmental vision differs from that of the project leaders, her options are to join some different project or to start a project of her own, not to create her own version of software already under development.

Informal sanctions are backed in part by the more formalized threat of legal action for copyright infringement. Attempts to capture open source projects in unapproved proprietary formats are potentially a violation of the license under which permission is granted to copy and adapt the code, and so they are a potential basis for a copyright suit. Such sanctions are of course the major deterrent to unauthorized use of the code by entities outside the open source community, but they also play a role in disciplining activity within the community, particularly where members of the community might be tempted to defect from the goals and values of the community. This function of community discipline potentially complicates attempts to employ copyleft strategies outside the context of software coding.

Open Source Patenting

The vision of open access to copyrighted works has not been limited to the open source community, and copyleft-style licensing, facilitated by the medium of electronic communications, has begun appearing outside the context of open source

software. For example, the Creative Commons project provides an online repository of off-the-shelf licenses, conforming to the laws of various nations from which copyright holders can choose to associate with their copyrighted works. A symbol attached to a given work refers the user of the work to a particular standardized license in the Creative Commons repository, where the terms of use for the work can be found. Thus, like open source licensing, Creative Commons licenses rely on the shrink-wrap licensing model, assuming that such standardized licenses form a viable contract and will be enforceable.

To encourage use of Creative Commons licensing for research publications and other copyrighted science materials, Creative Commons has recently turned its attention to scientific research via the Science Commons project (David, 2004). Creative Commons licenses may cover copyrightable aspects of databases, such as the selection and arrangement or output formats. At present they do not extend to uncopyrightable facts, to unoriginal arrangements of data, or to *sui generis* E.U. database rights. Of course, as with mass-market shrink-wrap licenses, Creative Commons or other open source licenses could be extended to cover use or alteration of otherwise uncopyrightable materials, using access to the database as the trigger for the license; access would be granted conditionally upon acceptance of the contract. Off-the-shelf forms for such licenses are not yet available via Creative Commons.

Nor do Creative Commons licenses address, at least at the present writing, the disposition of patent rights. Consequently, the use of such licenses is at best a partial solution to the disposition of intellectual property in electronic collaborations. In theory, there is no reason why the Creative Commons could not include off-the-shelf patent licensing contracts, or that a similar archive of such forms could not be generated.

Some research projects have attempted to generate their own version of open source patent licenses (Cockburn, 2005; Rai, 2005). For example, such licenses govern access to the genomic data assembled in the governmentally funded HapMap Project, which is building an informational haplotype map of human genetic variations. The database is freely available on the conditions that those accessing the data not file patent applications on information derived from the database and that those accessing the data share information only with others who have agreed to the same terms. Additionally, users of the database agree that any patents they obtain on uses derived from information in the database will be licensed on terms that allow others continued access to the information, limiting commercial use of the communal resource, and also incorporating the viral features of copyleft licensing.

Copyleft-style licensing has also been applied to physical materials, such as the biological materials made available via the Biological Innovation for Open Society (BIOS). The BIOS project is intended to make available biological research tools and techniques. While the project organizers are not adverse to users of these tools filing patents on discoveries made by use of the tools, they intend to preserve public access to the tools themselves. Patenting improvements or modifications that users might make to the basic tools might encumber those tools with proprietary claims,

threatening open access. Internet-based electronic resources offer information about the tools and their use and facilitate contact for physical transfer of the tools, but physical access is conditioned on agreement not to patent any improvements or modifications to the tools and to make any such modifications or improvements available on the same terms. No such restrictions are placed on products or discoveries generated by use of the tools; such products or discoveries can be patented without limitation.

However, the transfer of copyleft and related models from the copyright regime to that of patent presents several difficulties (Boettiger & Burk, 2004; Feldman, 2004). First, as detailed earlier, the nature of the exclusive rights granted by copyright and by patent are quite different. Copyright excludes unauthorized copying and related activities that are triggered by access to the protected work. Such access serves as the trigger or activating event for the copyleft license; copying or adapting the open source code opens the copyleft or adapters to a lawsuit unless the copying or adapting is done in accordance with the terms of the license. Patent rights exclude all uses of the claimed invention, even those conducted independently, without any access to the invention. In such cases, the infringing act would not serve to channel the infringer into compliance with the terms of the license, as there would be no knowledge, let alone manifestation of assent, to the license.

Second, the restrictions on further patenting that are incorporated into some “open biology” licenses may run afoul of the general public policy of the patent system. In the U.S., federal statutory and constitutional law encourages patenting, and licenses deterring patents may be preempted. Additionally, patents raise competition law considerations that are not necessarily present under copyright law. Certain types of patent licensing arrangements are subject to extra antitrust scrutiny, such as patent “pools,” in which participants cross license each others’ patents, patent “grant-backs,” which require licensing of technology developed with a patented tool back to the patent owner, and patent “reach-throughs,” which require payment of royalties to a patent owner for products developed with a patented tool. Patenting restrictions in open biology licenses resemble these types of arrangement—for example, requiring products developed with open source biology tools to be licensed back to others on an open source basis—and so may raise antitrust concerns.

But the greatest obstacle to movement of copyleft licenses into e-science may be the social disparity between open source and scientific research settings. Despite their apparent normative similarities, academic science as currently practiced has a different profile than that of the open source community. The scientific community is older and more institutionally invested, with an organizational structure not present in open source coding. For example, academic science has an effective hierarchical organization at the level of individual laboratories. Graduate and undergraduate training also contribute a distinct social sub-structure to science. Science is also arguably more diverse than open source coding; the norms of sociology are not necessarily the norms of physics or molecular biology. Academic science is also

heavily subsidized by governmental grants, with the result that funding agencies may have interests in the disposition of intellectual property. Other formal institutions, such as institutional ethics review boards and peer review journal publishers, may also play roles not contemplated by the open source licensing system.

Consequently, open source licenses used to promote open science may need to consider the role of these organizational structures, rather than blithely adopting licenses from open source programming or other open access projects. As one possibility, open science licenses might need to take into account scientific institutions with behavioral influence that could substitute for the strong informal norms found among open source programmers. For example, in the formulation and enforcement of human genome sequence publication rules discussed above, exclusion of violators from scientific journals played a key role in compliance. This suggests that open science licenses might include as their penalty clauses exclusion from peer review journals, or from peer review funding, or from participation in similar scientific institutions. Such contractual clauses could help match the structure of the license to the structure of the relevant research community.

Conclusion

The communal norms of science no longer shield research from the demands of intellectual property, and as a multijurisdictional and collaborative enterprise, e-science is potentially subject to a myriad of conflicting intellectual property regimes. Contractual systems drawn from open source software coding might be deployed to ameliorate some of the effects of conflicting intellectual property systems, but copyleft and related licensing systems were developed to maintain community discipline and demarcate the open source community from outside commercial influences. These licenses may require considerable adaptation before they can perform a similar function for the e-science research community. Further research and understanding of how the e-science community is being and will be structured are essential to assist in drafting licenses that meet the community's needs.

References

- Boettiger, S., & Burk, D. (2004). Open source patenting. *Journal of International Biotechnology Law*, 1(6), 221–231.
- Boyle, J. (2003). Enclosing the genome: What squabbles over genetic patents could teach us. In F. S. Keiff (Ed.), *Perspectives on Properties of the Human Genome Project* (pp. 97–122). San Diego, CA: Elsevier Academic Press.
- Burk, D. (1993). Patents in cyberspace: Territoriality and infringement on global computer networks. *Tulane Law Review*, 68, 1–67.
- Burk, D. (1994). Transborder intellectual property issues on the electronic frontier. *Stanford Law and Policy Review*, 5, 9–16.
- Burk, D. (1998). Virtual exit in the global information economy. *Chicago-Kent Law Review*, 73, 943–995.

- Burk, D. (2000). Intellectual property issues in electronic collaboration. In S. H. Koslow & M. F. Huerta (Eds.), *Electronic Collaboration in Science: Progress in Neuroinformatics, Vol. 2* (pp. 15–44). Mahwah, NJ: Lawrence Erlbaum Associates.
- Burk, D. (2005). Bioinformatics lessons from the open source movement. In H. Tavani (Ed.), *Ethics, Computing, and Genomics: Moral Controversies in Computational Genomics* (pp. 257–254). Sudbury, MA: Jones and Bartlett.
- Cockburn, I. (2005). State Street meets the human genome project: Intellectual property and bioinformatics. In R. W. Hahn (Ed.), *Intellectual Property Rights in Frontier Industries: Software and Biotechnology* (pp. 111–130). Washington D.C.: AEI-Brookings Joint Center.
- David, P. A. (2004). Can “open science” be protected from the evolving regime of IPR protections? *Journal of Institutional and Theoretical Economics*, *160*(1), 9–34.
- David, P., & Spence, M. (2003). Towards institutional infrastructures for e-science: The scope of the challenge. *Oxford Internet Institute, Research Report, 2*. Oxford: University of Oxford. Retrieved September 20, 2006 <http://129.3.20.41/eps/get/papers/0502/0502028.pdf>
- de Laat, P. (2001). Open source software: A new Mertonian ethos? In A. Vedder (Ed.), *Ethics and the Internet* (pp.33–48). Oxford: Intersentia.
- Eckersley, P., Egan, G. F., & Shun-ichi, A. (2003). Neuroscience data and tool sharing: A legal and policy framework for neuroinformatics. *Neuroinformatics*, *1*(2), 149–66.
- Eisenberg, R. (1987). Proprietary rights and the norms of science in biotechnology research. *Yale Law Journal*, *97*(2), 177–231.
- Feldman, R. (2004). The open source biotechnology movement: Is it patent misuse? *Minnesota Journal of Law, Science and Technology*, *6*(1), 117–167.
- Finholt, T. (2002). Collaboratories. *Annual Review of Information Science and Technology*, *36*(1), 73–107.
- Glasner, P. (1996). From community to “collaboratory?” The human genome mapping project and the changing culture of science. *Science and Public Policy*, *23*(2), 109–116.
- Goldsmith, J. (1998). Against cyberanarchy. *University of Chicago Law Review*, *65*(4), 1199–1250.
- Hey, T., & Trefethen, A. (2003). e-Science and its implications. *Philosophical Transactions of the Royal Society London*, *361*(1809), 1809–1825.
- Hinds, C., Jirotko, M., Rahman, M., D’Agostino, G., Meyer, C., Piper, T, et al. (2005, June 22-24). Ownership of intellectual property rights in medical data in collaborative computing environments. National Centre for e-Social Science, First International Conference on e-Social Science. Retrieved September 20, 2006 from http://www.ncess.ac.uk/events/conference/2005/papers/papers/ncess2005_paper_Hinds.pdf
- McGowan, D. (2001). Legal implications of open-source software. *University of Illinois Law Review*, *2001*(1), 241–304.
- Merges, R. P. (1996). Property rights theory and the commons: The case of scientific research. *Social Philosophy and Policy*, *13*(2), 145–167.
- Merton, R. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: University of Chicago Press.
- Nelson, R. (2004). The market economy and the scientific commons. *Research Policy*, *33*(3), 455–471.
- Newman, H. B., Ellisman, E. H., & Orcutt, J. A. (2003). Data-intensive e-science frontier research. *Communications of the ACM*, *46*(11), 69–75.

- Rai, A. K. (2000). Addressing the patent gold rush: The role of deference to PTO patent denials. *Washington University Journal of Law and Policy*, 2, 199–227.
- Rai, A. K. (2005). Open and collaborative research: A new model for biomedicine. In R. W. Hahn (Ed.), *Intellectual Property Rights in Frontier Industries: Software and Biotechnology* (131–158). Washington D.C.: AEI-Brookings Joint Center.
- Reichmann, J. H., & Uhlir, P. F. (2003). A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment. *Law and Contemporary Problems*, 66(1 & 2), 315–462.
- Toumi, I. (2000). Internet, innovation, and open source: Actors in the network. *First Monday*, 6(1). Retrieved September 20, 2006 from http://www.firstmonday.org/issues/issue6_1/toumi/index.html

About the Author

Dan L. Burk holds the Oppenheimer, Wolff & Donnelly Professorship in Law at the University of Minnesota, where he teaches courses in patent law, copyright, and electronic commerce. An internationally prominent authority on issues related to high technology, he is the author of numerous papers on the legal and societal impact of new technologies, including articles on scientific misconduct, on the regulation of biotechnology, and on the intellectual property implications of global computer networks.

Address: University of Minnesota Law School, 229 19th Avenue South, Minneapolis, MN 55455 USA