Paper presented at the conference

INFORMATION, INTELLECTUAL PROPERTY,AND ECONOMIC WELFARE
Turin, Italy May 15-16 2006 , Fondazione Luigi Einaudi

URL: http://polis.unipmn.it/seminari/information/information_file/frame.htm

Patents and Data Sharing in Public Science

Rebecca S. Eisenberg

All of us who write about patent law owe a great debt to Nelson's *The Simple Economics of Basic Scientific Research*[1] and Arrow's *Economic Welfare and the Allocation of Resources for Invention*.[2] Most of the basic analytical tools that we still use to understand the patent system are presented in these two papers with great clarity and common sense. These papers show economics at its best, helping to clarify intuitions and to examine them rigorously in order to understand vexing social problems.

The essential problem for both Nelson and Arrow is how best to get information produced and disseminated. Private markets alone are likely to underinvest in the production of information because the outcome of research is uncertain and because it is difficult to capture its benefits privately. To the extent that the benefits of information production are privately appropriated, its social value is reduced because access is priced above marginal cost. These problems are greater for basic research than for applied research, arguing for government sponsorship of basic research in institutions such as universities that are committed to dissemination of research results. These early sketches of the economic problems that lie behind IP law have held up remarkably well over the past 45 years, despite significant ongoing changes in both law and technology.

Nonetheless, two changes in the intervening years–one primarily legal and the other primarily technological–represent a sufficient departure from the prior state of the world as to call for ongoing reexamination of some old assumptions about the economics of R&D: first, an increase over time in the appropriability (and appropriation) of basic research results through the patent system; second, the revolution in information technology and information networks, which has had reverberations outside the law of intellectual property in the evolving norms of the scientific community regarding data-sharing. I examine these changes with particular attention to their impact on biomedical research, an area in which the patent system is generally thought to be particularly important to the incentives of innovators.

The Appropriation Shift

Writing in 1959, Nelson observed that "basic research efforts are likely to generate substantial external economies" because "significant advances in scientific knowledge are often not directly and immediately applicable to the solutions of practical problems and hence do not quickly result in patents."[3] The intervening decades have witnessed a significant shift in favor of patenting the results of basic scientific research, as has been

[1] Richard R. Nelson, "The Simple Economics of Basic Scientific Research" *J. Pol. Econ*. 67:297-306 (1959).
[2] Kenneth J. Arrow, "Economic Welfare and the Allocation of Resources for Invention" in *The Rate and Direction of Inventive Activity* (Princeton 1962).
[3] Nelson at 302.

much remarked upon in in both the legal literature[4] and the economics literature,[5] and even in the popular press.[6]

A number of factors are generally credited with contributing to this perceived trend. The Bayh-Dole Act of 1980[7] locked in place for all federal research sponsors a policy already followed by some sponsors of allowing contractors and grantees to retain ownership of inventions made with federal funds.[8] Around the same time, advances in molecular biology suggested a (perhaps deceptively) clear path from the basic research laboratory to the biopharmaceutical product market,[9] motivating universities to patent the discoveries of their biomedical scientists and to license them to private firms for commercial development. A new biotechnology industry emerged with for-profit firms trying to stake out market niches performing research that lay somewhere between the basic research that was occurring in academic laboratories and the end-product development that was occurring in established pharmaceutical firms. Meanwhile, the decision of the U.S. Supreme Court to uphold a patent claim on a genetically-engineered microorganism in *Diamond v. Chakrabarty*,[10] followed by the creation of the pro-patent Court of Appeals for the Federal Circuit,[11] promised a favorable legal environment for appropriation of emerging technologies.[12] In this environment, upstream patent-seekers have become ever more creative in their claiming strategies, devising new ways of capturing the expected value that their early-stage discoveries may contribute to future product development through patent claims that dominate potential future products.[13]

---

[4] Eisenberg, R.S., "Public Research and Private Development: Patents and Technology Transfer in Government-Sponsored Research" *Virginia Law Review* 82: 1663-1727 (1996); Margo A. Bagley, "Academic Discourse and Proprietary Rights: Putting Patents in Their Proper Place" *B.C. L. Rev* 47:217-74 (2006).

[5] See, e.g., David C. Mowery, Richard R. Nelson, Bhaven N. Sampat & Arvids A. Ziedonis, Ivory Tower and Industrial Innovation: University-Industry Technology Transfer Before and After the Bayh-Dole Act (Stanford 2004); Adam B. Jaffe & Josh Lerner, Innovation and Its Discontents: How Our Broken Patent System Is Endangering Innovation & Progress, and What to Do About It (2003)

[6] See Clifton Leaf, "The Law of Unintended Consequences" *Fortune,* Sept. 19, 2005, at 250.

[7] Act of Dec. 12, 1980, Pub. L. No. 96-517, Section 6(a), 94 Stat. 3015, 3019-28 (1980) (codified as amended at 35 U.S.C. Sections 200-212 (1994)).

[8] See Rebecca S. Eisenberg, Public Research and Private Development: *Patents and Technology Transfer in Government-Sponsored Research*, 82 VA. L. REV. 1663-1727 (1996).

[9] *Id.* at 289.

[10] 447 U.S. 303 (1980).

[11] Federal Courts Improvement Act of 1982, Pub.L.No. 97-164, 96 Stat. 25 (codified as amended at 28 U.S.C. sections 41 et seq. (1982)).

[12] Kortum, S., and J. Lerner (1999) "What Is Behind the Recent Surge in Patenting?" *Research Policy* 28:1-22.

[13] For example, fundamental research to explain a biochemical pathway may support patent claims to methods of treatment that act upon that pathway, giving rise to infringement claims against later-developed drugs that act upon the pathway. Although the U.S. Court of Appeals for the Federal Circuit has used the written description requirement to invalidate one such patent, see University of Rochester v. G.D. Searle, 358 F.3d 916, rehearing en banc denied, 375 F.3d 1303 (Fed. Cir. 2004) (affirming judgment of invalidity of patent claiming method of treatment involving selective inhibition of PGHS-2 enzyme but not PGHS-1 enzyme), a more recent jury decision in a federal District Court in Boston has come out the other way. See Peter Loftus, Lilly Loses Patent Fight With Ariad, Wall St. J. (May 4, 2006) (available at http://online.wsj).

Nelson and Arrow might have predicted that greater appropriability would bring greater private investment in R&D.[14] Some evidence appears to bear this out, although it is inconclusive. Industry funding for R&D in the U.S. has exceeded federal government funding since around 1980,[15] although industry's share of funding for the elusive category of "basic research" remains stubbornly low.[16] In biomedical research, industry currently supplies approximately 57% of all funding for biomedical research, while federal, state, and local government funding combined total approximately 36%.[17] It is unclear how far changes in patent law, including the Bayh-Dole Act, may take credit for these temporal shifts in the relative shares of public and private investment in R&D.

In some research areas, it has become hard to draw bright lines between the types of research occurring in private laboratories and the types of research occurring in public laboratories. The convergence in public and private research goals became quite dramatic in the latter stages of the Human Genome Project, as publicly-funded academic investigators raced against researchers in the private firm Celera to complete the first DNA sequence of the human genome.[18]

Nelson in 1959 and Arrow in 1962 may not have predicted the current extent of private appropriability of discoveries arising from basic biomedical research. But to their great credit, they did give us tools for thinking about it. Although both Nelson and Arrow lamented the inadequacy of private incentives to perform fundamental research in the face of uncertainty and limits on appropriability, neither one called for fortifying patent protection as a way of fixing the problem. Instead, both recognized that, given the characteristics of information as a public good, more complete appropriation would lead to dynamic inefficiencies through restrictions on utilization.[19] Careful readers of these two foundational papers therefore cannot leap to the conclusion that the economic problems they described are receding in importance in the face of a more robust patent system than they envisioned. Instead, we must examine the tradeoff between two effects of private appropriations: (1) greater private incentives for research performance; and (2) underutilization of research results, and ask whether we have drawn the correct balance.

---

[14] Arrow at 617; Nelson at 302-304.
[15] National Science Foundation, Science & Engineering Indicators 2006, App. Table 4-5, posted on the internet at http://www.nsf.gov/statistics/seind06/append/c4/at04-05.pdf (visited 7/17/2006
[16] *Id.* App. Table 4-9, at http://www.nsf.gov/statistics/seind06/append/c4/at04-09.pdf (visited at 7/18/2006)
[17] Hamilton Moses et al., "Financial Anatomy of Biomedical Research" *Journal of the American Medical Ass'n* 294:1333-42 (2005)
[18] Eisenberg, R.S. & Nelson, R.R., "Public vs. proprietary science: a fruitful tension?" *Daedalus* (Spring 2002) 89-101.
[19] Arrow at 616-17 ("As we have seen, information is a commodity with peculiar attributes, particularly embarrassing for the achievement of optimal allocation. In the first place, any information obtained, say a new method of production, should, from the welfare point of view, be available free of charge …. This insures optimal utilization of the information but of course provides no incentive for investment in research…. The first problem, then, is that in a free enterprise economy the profitability of invention requires a nonoptimal allocation of resources."); Nelson at 306 ("The marginal social cost of using knowledge that already exists is zero. For maximum static economic efficiency, knowledge should be administered as a common pool, with free access to all who can use the knowledge. But, if scientific knowledge is thus administered, the incentives of private firms to create new knowledge will be reduced.")

As Nelson observed in 1959,

> Often … new knowledge is of greatest value as a key input of other research projects which, in turn, may yield results of practical and patentable value. For this reason scientists have long argued for free and wide communication of research results, and for this reason natural "laws" and facts are not patentable.[20]

By contrast, the premise of the Bayh-Dole Act is that licenses to use patented upstream research discoveries to develop downstream products will promote product development by protecting licensees from free riders. This may be a plausible assumption in some circumstances, but it also stands to reason that sometimes patents on research discoveries will diminish incentives for further R&D by making it more costly and by allowing upstream patent owners to force product developers to share expected rents.

Universities and their licensees have become increasingly aggressive about suing product-developing firms for patent infringement,[21] sometimes recovering substantial damage awards or settlements.[22] Such awards and settlements seem likely to dampen the expected profits from product development in areas dominated by patents on early research discoveries.

Academic institutions have so far had mixed success in using their patents on "upstream" research discoveries to appropriate value from "downstream" product developers. A major challenge, foreseen by Arrow, is the difficulty of devising patent claims that would permit appropriation of the contribution that past discoveries make to future inventions.[23]

A recurring motif that is currently working its way through the U.S. courts involves procuring patents on methods of treatment that act on biochemical pathways that have been explained in the course academic research and enforcing those patents against firms that develop drugs designed to act in those pathways.

This strategy was unsuccessful in *University of Rochester v. G.D. Searle.*[24] The patent at issue in that case grew out of research in the early 1990s by scientists at the University of Rochester. The researchers figured out that two related enzymes, PGHS-1 and PGHS-2, both inhibited by then-available nonsteroidal anti-inflammatory drugs (NSAIDs), had different functions. Finding that PGHS-1 protects the stomach lining, while PGHS-2 promotes inflammation, the scientists surmised that a drug that selectively inhibits only PGHS-2 would provide relief from pain and inflammation without the gastrointestinal side effects often associated with use of NSAIDs. They did not themselves identify any such drug, but drafted patent claims that would be infringed by the use of such products.

---

[20] Nelson at 302.

[21] See, e.g., Regents v. Eli Lilly, 119 F.3d 1559 (Fed. Cir. 1997); University of Rochester v. G.D. Searle, 375 F.3d 1303 (Fed. Cir. 2004); Ariad Pharmaceuticals v. Eli Lilly, 2004 U.S. Dist. LEXIS 3170 (D. Mass. 2004).

[22] See, e.g., Univ. of Calif. Press release, "UC, Monsanto reach $100 million settlement in growth hormone patent case" (Feb. 27, 2006), posted at <http://www.universityofcalifornia.edu/news/2006/feb27b.html>

[23] Arrow, *Economic Welfare and the Allocation of Resources for Invention*, at 617-18.

[24] 375 F.3d 1303 (Fed. Cir 2004)

They then sued pharmaceutical firms that had meanwhile developed selective PGHS-2 inhibitors (such as Celebrex and Vioxx) for infringement of their patented "method for selectively inhibiting PGHS-2 activity in a human host" in which "the activity of PGHS-1 is not inhibited." The Federal District Court held the patent invalid for failure to provide a "written description" of the invention and for failure to provide a disclosure that was sufficient to enable others to make and use the invention as required by § 112 of the U.S. Patent Act,[25] and a Federal Circuit panel affirmed on grounds of inadequate "written description" without reaching the "enablement" issue.[26]

A different set of patent owners have enjoyed greater success with a similar strategy in recent litigation involving patents growing out of research at Harvard, MIT, and the Whitehead Institute on the transcription factor NF-kB. NF-kB appears to be involved in activating and deactivating more than 175 genes and plays a role in many disease pathways.[27] The patent, which broadly claims "[a] method for inhibiting expression, in a eukaryotic cell, of a gene whose transcription is regulated by NF-.kappa.KB,"[28] potentially dominates many drugs, including the Eli Lilly drugs Evista and Xigris. In defending against an infringement action, Eli Lilly argued that the patent was invalid on two grounds: first, that the two products in suit, both of which were developed and patented before the inventive activities that led to the patent claims in issue, rendered the broad claims to the NF-kB invention invalid for lack of novelty; and second, that the patent merely expresses a desired result without disclosing how to achieve that result and is thus invalid for lack of enablement. The trial court rejected Lilly's motion for summary judgment in a 2003 decision, prior to the Federal Circuit's decision in the *University of Rochester* case, stating that the novelty issue had not been adequately developed at that stage of the litigation and noting that the enablement issues "is a serious and troubling one."[29] A jury in Boston recently found the patent valid and infringed and returned a verdict against Eli Lilly for $65.2 million plus a 2.3% royalty on future product sales.[30] The verdict may yet be set aside, either by the trial judge or by the Federal Circuit,[31] but if it stands, it suggests that patents on upstream discoveries have come very far as a mechanism for appropriating the value that basic research contributes to downstream product development.

To the extent that such patents are enforceable, they may force drug-developing firms to share the rents that they earn from their patented products with upstream research institutions such as universities. Such forced rent-sharing is in tension with the goal of the Bayh-Dole Act to promote product development. Whether it promotes social welfare depends in part on whether we are more concerned with the adequacy of incentives of

---

[25] 35 U.S.C. §112.

[26] Univ. of Rochester v. G.D. Searle & Co., 358 F.3d 916, *rehearing en banc denied*, 375 F.3d 1303 (Fed. Cir. 2004).

[27] Ken Garber, *Biomedical Patents: Broad Patent Faces Narrow Odds in Court Battle*, 311 Science 1855-57 (2006); Patrick A. Baeuerle & David Baltimore, *NF-kB: Ten Years After*, 87 Cell 13-20 (1996)

[28] U.S. Pat. No. 6,410,516, issued June 25, 2002, claim 1.

[29] Ariad Pharmaceuticals v. Eli Lilly, 2003 U.S. Dist. LEXIS 8030 (D. Mass. 2003)

[30] Andrew Pollack, *Lilly Loses Patent Case to Ariad*, N.Y. Times (May 5, 2006).

[31] Eli Lilly, Press Release, *Jury Issues Verdict in Ariad v. Lilly Patent Infringement Suit* (May 4, 2006), posted at http://newsroom.lilly.com/ReleaseDetail.cfm?ReleaseID=195311.

upstream research performers (who rely heavily on public funding) or of downstream product developers (who rely on funding from private investors). To the extent that such rent-sharing is achieved through the creation of new exclusive rights that limit dissemination of information that would otherwise be more broadly disseminated and used, it imposes welfare losses that may well exceed any welfare gains from enhanced incentives for upstream research.

It is less clear that patents are dampening incentives to engage in further basic research in the nonprofit sector. Although there are some fields, such as agriculture, in which academic scientists experience patents as a significant obstacle to their research,[32] some data suggest that many academic scientists simply ignore patents, perhaps assuming that the patent laws do not apply to their research activities. The Federal Circuit squarely held that basic research in universities is not exempt from infringement liability in its 2002 decision in *Madey v. Duke*,[33] causing academic institutions to become more concerned about infringement liability. Yet one recent empirical study found that, even post-*Madey,* patents have had relatively little impact on the work of academic scientists, most of whom are largely oblivious to the patents that they may be infringing.[34]

Although no firm conclusions are possible on the basis of such limited data, a plausible story may be told that would make sense of these observations. Following the *Madey* decision, many universities took steps to alert their faculty to the importance of avoiding infringement as a general matter. But the response of individual faculty members to such general admonitions is likely to be uneven at best. Universities have little incentive to identify acts of infringement by faculty members, given that knowing violations of patent rights could make them liable for higher damage remedies for willful infringement if they get sued.[35] Patent holders occasionally send demand letters to universities notifying them that they are infringing their patents and asking them to enter into license agreements,[36] but infringement actions against universities remain highly unusual.[37] In the face of this apparent reticence to sue, it is not surprising that patents in and of themselves might only rarely pose an obstacle to the research plans of academic scientists. More empirical research would be useful to test this controversial hypothesis.

Even as patent lawyers get better at drafting patent claims that will dominate the output of follow-on research, they face a persistent challenge in enforcing the rights that they obtain. A patent is a license to sue, and the burden of inertia rests heavily on the patent holder. Infringement litigation is costly and its outcome is uncertain. Unless there is a

---

[32] See Richard C. Atkinson et al., *Public Sector Collaboration for Agricultural IP Management*, 301 Science 174-75 (2003).
[33] 307 F.3d 1351 (Oct. 3, 2002).
[34] John P. Walsh et al., *View from the Bench: Patents and Material Transfers*, 309 Science 2002-03 (2005).
[35] See Crystal Semiconductor v. TriTech Microelectronics Int'l, 246 F.3d 1336, 1351 (Fed. Cir. 2001) ("When an infringer has actual notice of a patentee's rights, the infringer has an affirmative duty of due care to avoid infringement.")
[36] See Sam Jaffe, *Ongoing Battle over Transgenic Mice*, The Scientist (July 19, 2004) at 46-47.
[37] Academic institutions are sometimes sued for infringement along with their private firm collaborators. E.g., *Merck v. Integra*, in which the Scripps Institute was joined as a defendant.

conspicuously lucrative endpoint for the infringing activity in the near term, patent holders may conclude that the costs of enforcement do not justify the potential gains. Litigation may therefore be less likely when infringement occurs in purely academic research than when it occurs in research with commercial implications, although some patent owners may be motivated to sue even over purely academic research.[38]

Academic scientists are much more likely to report that they have been stymied in their requests for access to tangible materials and data than that they have been stymied by patents.[39] There is an obvious reason why access to data and materials is more likely to be problematic than access to patented inventions: researchers need the cooperation of those who control access to data and materials before they can get their hands on them. The burden of inertia is thus on the user. The holder of the resources need not bring an infringement action in order to force the user to pay, but can sit back and wait for the user to seek access and then bargain over terms. By contrast, if only patents are at stake, users can simply ignore a potential claim of infringement and go ahead with their research plans, leaving the burden of inertia on the patent owner to discover the activity and try to stop it. Indeed, when researchers cite patents as obstacles to scientific research, often, upon closer examination, the obstacle consists of restrictive terms of access to materials (such as knock-out mice) that are covered by patents.[40] Patents on research methods or other inventions that can be readily duplicated in the laboratory without first obtaining permission from the owner are less likely to present an obstacle to research, because many researchers simply ignore them and get away with it. The obstacle that academic researchers take note of is less likely to be a patent than a restriction on access to something that is costly to duplicate without a license.

Of course, this is not to say that patents alone never pose an obstacle to research. If infringing activities appear to be commercially significant, patent owners are more likely to find it worthwhile to overcome the burden of inertia and enforce their rights. For this reason, product-developing firms are more likely to be targets of infringement actions than academic institutions, and such firms are more likely to find patents to be an obstacle to their work.

A number of recent studies have considered the impact of patents on academic research. More empirical research might help clarify the extent to which patents are restricting biomedical research and product development within private firms.

---

[38] See, e.g., *Madey v. Duke,* 307 F.3d 1351 (infringement action filed against university by scientist who had been removed from positions as laboratory director and principal investigator on grant).

[39] Eric G. Campbell et al., *Data withholding in academic medicine: characteristics of faculty denied access to research results and biomaterials*, 29 Research Policy 303-312 (2000); Eric G. Campbell et al., *Data Withholding in Academic Genetics: Evidence From a National Survey*, 287 JAMA 473-80 (2002); John P. Walsh, et al., *View from the Bench*, *supra*; Rebecca S. Eisenberg, *Bargaining Over the Transfer of Proprietary Research Tools: Is This Market Failing or Emerging?*, in R. Dreyfuss et al. eds., Expanding the Boundaries of Intellectual Property: Innovation Policy for the Knowledge Society (Oxford 2001) at 223-49.

[40] An example of this is the Harvard-owned patents on cre-lox P technology for creating knock-out mice, licensed exclusively to DuPont. See National Research Council, *Finding the Path: Issues of Access to Research Resources* 10-11 (Nat'l Acad. Press 1999)

Information Technology and Data-Sharing

In 1959, Nelson made a powerful case for government funding of basic research in universities in part on the ground that academic researchers were committed to "full and free dissemination of research results."[41] But although research scientists may embrace free dissemination as a norm of aspiration, there is considerable evidence to suggest that over the years this norm has often been observed in the breach. Data-sharing has been a recurring focus of struggle within the biomedical research community as improvements in information technology and digital networks have expanded the ways that data can be produced, disseminated, and used.[42] The experience of this community sheds an interesting light on the norms and incentives of public sector scientists and provides a more complex picture of how they interact with the norms and incentives of private firms and other patent-seeking innovators.

Information technology makes it easier to share data in publicly accessible archives that aggregate data from multiple sources, facilitating observations that would otherwise be impossible. But data disclosure poses a dilemma for scientists. Data have long been the stock in trade or working scientists, lending credibility to their claims while highlighting new questions that are worthy of future research funding. Some disclosure is necessary in order to claim these benefits through publication, but data disclosure may also benefit one's research competitor's tomorrow and beyond.[43] Scientists who share their data promptly and freely may find themselves at a competitive disadvantage relative to free riders in the race to make and publish future observations, and thereby to earn further recognition and funding. The emergence of bioinformatics experts who are particularly adept at analysis of data gathered by others makes data-gatherers particularly vulnerable to being "scooped" if they make their data freely available in the public domain. The possibility of commercial gain further raises the competitive stakes. As information technology has advanced, and as commercial interests in biomedical research have grown, this dilemma has become more pronounced.

Intellectual property law has so far played a relatively small role in debates about data-sharing. Data are generally considered ineligible for either copyright or patent protection.[44] The Bayh-Dole Act,[45] which governs patent rights in the results of

---

[41] Nelson at 302, 305.

[42] See, e.g., National Research Council, Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences (Nat'l Acad. Press 2003). See generally National Research Council, Bits of Power: Issues in Global Access to Scientific Data (Nat'l Acad. Press 1997).

[43] See, e.g., Eliot Marshall, *DNA Sequencer Protests Being Scooped With His Own Data*, 295 Science 1206-07 (2002).

[44] For a review of the limits on copyright protection of data with citations to the relevant cases and literature, see J.H. Reichman & Paul F. Uhlir, *A Contractually Reconstructed Research Comnmons for Scientific Data in a Highly Protectionist Intellectual Property Environment*, 66 L. & Contemp. Prob. 315, 336-341 (2003). For a review of the limits on patent protection of data, see U.S. Pat. & Trademark Off., Interim Guidelines for Examination of Patent Applications for Patent Subject Matter Eligibility (2005), posted at <http://www.uspto.gov/web/offices/com/sol/og/2005/week47/patgupa.htm>.

federally-sponsored research, has nothing to say about the dissemination of unpatentable data. Although *sui generis* database protection has been enacted in Europe[46] and proposed in the U.S.,[47] it has not yet been passed into law in the U.S. and appears to have had little impact in Europe.

Indeed, both copyright law and patent law treat the informational content of writings and inventions as a spillover benefit for the public, while limiting the exclusionary rights of creators to something else: an original expression in the case of copyright;[48] and a useful invention in the case of patent.[49] Both legal regimes nonetheless tend to promote information dissemination by offering exclusive rights that survive even after disclosure. Moreover, the patent system requires disclosure as a precondition for securing rights.[50]

On one reading, the failure to protect information under patent law and copyright law shows that information gets no respect. This is the sense that emerges from reading copyright cases like *Feist.*[51] In this story copyright law treats information as a mere byproduct of efforts that deserve protection only insofar as they yield something else that is more creative. Contemporary critics charge that IP law has failed to appreciate the importance of information as an artifact of human ingenuity with value in its own right, and as this value grows and becomes more vulnerable to misappropriation with

---

[45] Act of Dec. 12, 1980, Pub. L. No. 96-517, 94 Stat. 3015 (1980) (codified as amended at 35 U.S.C. §§ 200-212).

[46] Council Directive 96/9 of 11 March 1996 on the Legal Protection of Databases, 1996 O.J. (L 77).

[47] For a review of these proposals from the perspective of the scientific community, see J.H. Reichman & Paul F. Uhlir, *Database Protection at the Crossroads: Recent Developments and Their Impact on Science and Technology* Berkeley Tech. L.J. 793 (1999).

[48] Cf. *Feist Publications v. Rural Telephone Service Co.*, holding that an alphabetized list of names and phone numbers lacked the minimum originality necessary for copyright protection, even though considerable effort may have gone into creating it.

[49] Patentable subject matter is limited by statute to any new and useful process, machine, manufacture, or composition of matter, 35 U.S.C. § 101, all generally understood to be limited to something tangible, as distinguished from data or information. The subject matter boundaries of the patent system have been falling away in recent judicial decisions in the face of creative claiming strategies for new technologies, particularly information technology. See, e.g., State Street Bank & Trust v. Signature Financial Group, 149 F.3d 1368 (Fed. Cir. 1998), *cert. denied*, 119 S. ct. 851 (1999); AT&T Corp. v. Excel Communications, Inc., 172 F.3d 1352 (1999). On the other hand, the Supreme Court recently granted certiorari in the case of *Laboratory Corporation of America Holdings v. Metabolite Laboratories Inc.*, 370 F.3d 1354 (Fed. Cir. 2004), *cert. granted,* 126 S. Ct. 543 (2005), *vacated, reconsidered, and cert. granted,* 126 S. Ct. 601 (2005), limiting the scope of its review to the question of patentable subject matter. This decision could potentially alter the trend toward more expansive patent eligibility.

[50] 35 U.S. Code § 112.

[51] "The sine qua non of copyright is originality. To qualify for copyright protection, a work must be original to the author. [] Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity. [] To be sure, the requisite level of creativity is extremely low; even a slight amount will suffice. The vast majority of works make the grade quite easily, as they possess some creative spark, "no matter how crude, humble or obvious" it might be. [] ... [F]acts do not owe their origin to an act of authorship. The distinction is one between creation and discovery: the first person to find and report a particular fact has not created the fact; he or she has merely discovered its existence. ... [O]ne who discovers a fact is not its "maker" or "originator." [] "The discoverer merely finds and records."" 499 U.S. 340, 345-57 (1991).

expanding capabilities of IT, this limitation on legal rights is becoming more anomalous.[52]

From another perspective, the failure to protect information reflects a reverence for information as something that is too important to society to permit it to be monopolized. This is the sense that emerges from reading cases about disclosure in the patent system, in which courts treat the informational content of patent applications as the public's *quid pro quo* that justifies the issuance of patents.[53] In this story, disclosure of unprotected information is not an incidental byproduct of a process that aims to motivate something more worthwhile, but the whole point of the patent system. In other words, we promote disclosure of precious information by rewarding disclosure with exclusionary rights in something else.

The absence of federal IP rights in data under U.S. law is by no means a guarantee that data will be shared freely in the public domain. Quite the contrary, in the absence of statutory protection like a patent or copyright that survives beyond disclosure, a standard strategy for preserving the commercial value of data and databases has been secrecy, or more accurately, restricted access. Some valuable databases are used only internally within a firm, protected under the law of trade secrecy, or made available only to paying subscribers under the terms of database access agreements, protected under the law of contracts or through technological restrictions on access to a website

These strategies allow database owners to exclude free riders, and perhaps thereby to capture enough value to justify creating the database. Licensed access to proprietary databases may promote efficiency by making database creation profitable and by providing users with access to information that would otherwise be unavailable to them at reasonable cost. Moreover, the interactions that are necessary to gain access may bring about useful scientific interactions that would not otherwise occur.

But readers of Nelson and Arrow will recognize that restrictions on access entail some degree of social waste because they limit dissemination of information that would have greater social value if it were made more broadly available, and that could be made freely available at minimal additional cost. Restricting access may also lead to socially wasteful

---

[52] See, e.g., Jane C. Ginsburg, *U.S. Initiatives to Protect Works of Low Authorship*, in R. Dreyfuss et al., eds, Expanding the Boundaries of Intellectual Property: Innovation Policy for the Knowledge Society (Oxford 2001)

[53] See, e.g., Kewanee Oil v. Bicron, 416 U.S. 470, 481 (1974)("When a patent is granted and the information contained in it is circulated to the general public and those especially skilled in the trade, such additions to the general store of knowledge are of such importance to the public weal that the Federal Government is willing to pay the high price of 17 years of exclusive use for its disclosure, which disclosure, it is assumed, will stimulate ideas and the eventual development of further significant advances in the art."); Bonito Boats v. Thundercraft, 489 U.S. 141, 151 (1989) ("[T]he ultimate goal of the patent system is to bring new ideas and technologies into the public domain through disclosure. State law protection for ideas and designs whose disclosure has already been induced by market rewards may conflict with the very purpose of the patent laws by decreasing the range of ideas available as the building blocks of further innovation.") United States v. Dubilier Condenser Corp. 289 U.S. 178, 186-87 (1933) ("[The inventor] may keep his invention secret and reap its fruits indefinitely. In consideration of its disclosure and the consequent benefit to the community, the patent is granted.").

duplication of effort if competitors are forced to recreate similar databases for their own use. The need to negotiate for licensed access to databases is particularly problematic for projects that involve aggregating data from multiple sources, as the transaction costs could mount quickly and perhaps become prohibitive.[54] As information technology advances in ways that facilitate meta-analysis of aggregated data, these transaction costs loom larger as an obstacle to full utilization of proprietary data. Nonetheless, trade secrecy and licensed access serve important functions in encouraging firms to invest in the creation of databases.

The case for trade secrecy and restricted access is weaker for information generated at public expense. Public funding mitigates concerns about the adequacy of incentives to generate information, and makes the social waste inherent in secrecy more troubling. In recent years the U.S. National Institutes of Health (NIH) have required researchers applying for more than $500,000 in funding to submit a plan for data sharing in their grant applications.[55] NIH cites a compelling list of arguments in support of data-sharing, including reinforcing open scientific inquiry, facilitating new research, encouraging diversity of analysis and opinions, enabling the exploration of topics not envisioned by the original investigators, and permitting the creation of new data sets that combine data from different sources.[56] Data-sharing can also provide a valuable check on overclaiming, both in the scientific and political arenas.

Meanwhile, the scientific community has sought to clarify its data-sharing norms and to figure out how to implement them. One important focus of debate has been the extent of data disclosure that should accompany scientific publication.[57] Although disclosure of research results is the essence of scientific publications, typically print publications only reveal data in summary form, leaving authors with substantial control over access to the underlying raw data. In an earlier era such summary disclosures may have been the best that could be achieved as a practical matter given scarcities of space associated with print media. But with the growth of the internet and information technology, it is now a simple matter to make vast data sets available over the internet at minimal cost. Yet a recent survey found that less than half of the most frequently cited journals in the life sciences and medicine had policies requiring deposit of data associated with published articles.[58]

Debate within the scientific community over the disclosure obligations associated with publication reached a fevered pitch with the publication in the prestigious *Science* magazine of an article announcing the completion of the human genome sequence by

---

[54] See Michael S. Heller & Rebecca S. Eisenberg, "Can Patents Deter Innovation? The Anticommons in Biomedical Research" *Science* 280:698-701.
[55] NIH Data Sharing Policy and Implementation Guidance (March 5, 2003), posted at <http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm>.
[56] NIH Data Sharing Policy and Implementation Guidance (March 5, 2003), posted at <http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm>.
[57] See, e.g., National Research Council, Finding the Path: Issues of Access to Research Resources (1999); NATIONAL RESEARCH COUNCIL, SHARING PUBLICATION-RELATED DATA AND MATERIALS: RESPONSIBILITIES OF AUTHORSHIP IN THE LIFE SCIENCES (Nat'l Acad. Press 2003)(hereinafter SHARING PUBLICATION-RELATED DATA AND MATERIALS).
[58] *Id.* at 33 Table 2-1.

scientists at the private firm Celera.[59] Although Celera made its human genome sequence available free of charge from its own website, access was restricted along certain dimensions, including quantitative limitations on the amount of data that could be downloaded, a prohibition on redistribution, and additional limitations on commercial users.[60]

The National Research Council took charge of the ensuing debate by forming a Committee on Responsibilities of Authorship in the Biological Sciences to examine the topic of sharing published data and materials. After an open meeting at which diverse viewpoints were expressed,[61]that Committee issued a report that called upon authors to include in their publications or otherwise make freely available "the data, algorithms, or other information that is central or integral to the publication – that is, whatever is necessary to support the major claims of the paper and would enable one skilled in the art to verify or replicate the claims."[62] The report specifically condemned the terms of access to the Celera human genome sequence data as "not consistent with the principles laid out in this report," noting that it permitted only "static access" for purposes of validation and not "dynamic access" for use in further research.[63]

Research projects that have as their aim the creation of large data sets over an extended period of time have presented special challenges for the implementation of data-sharing norms. The usual trigger for disclosure in academic research is publication of research results. But it may take years to complete large data sets, such as the sequence of the human genome, to the point that the accumulated total is ripe for publication in a prestigious journal, raising the possibility of delaying disclosure until long after data are generated.

A series of international collaborative research efforts in genomics that specifically aim to create resources for the broad scientific community have sought to require data-sharing prior to publication.[64] As expressed in the Bayh-Dole Act, the legal policy of the U.S. government is generally to favor the patenting of government-sponsored research results by contractors and grantees. But the value of free access to scientific information has had considerable rhetorical force in justifications for public funding of certain "big science" biomedical research projects, such as the Human Genome Project, especially when public and private research overlap. When the private firm Celera announced plans to complete the sequence of the human genome in private laboratories before the expected completion of the publicly-funded Human Genome Project, the public sponsors argued for continued

---

[59] J. Craig Venter et al., *The Sequence of the Human Genome*, 291 SCIENCE 1304-51 (2001).
[60] *Accessing the Celera Human Genome Sequence Data*, posted at <http://www.sciencemag.org/feature/data/announcement/gsp.dtl>.
[61] Eliot Marshall, *Clear-Cut Publication Rules Prove Elusive*, 295 SCIENCE 1625 (2002).
[62] SHARING PUBLICATION-RELATED DATA AND MATERIALS AT 5
[63] *Id.* at 48 Box 3-2.
[64] See, e.g., Summary of principles agreed at the International Strategy Meeting on Human Genome Sequencing, Bermuda, 25th-28th February 1996, <http://www.gene.ucl.ac.uk/hugo/bermuda.htm>; Sharing Data from Large-Scale Biological Research Projects: A System of Tripartite Responsibility, Report of a meeting organized by the Wellcome Trust and held on 14-15 January 2003 at Fort Lauderdale, USA, <http://www.wellcome.ac.uk/assets/wtd003207.pdf>

funding of an accelerated public sector effort. They stressed the importance of unrestricted access to the human genome sequence, ensured by requiring grantees to deposit newly identified sequence in the public Genbank database within twenty-four hours,[65] and questioned whether Celera would make as much information available as quickly to the scientific community.

Another strategic consideration that appears to have played a role in the development of disclosure rules for the Human Genome Project is the creation of patent-defeating "prior art." Patent law defines the scope of the prior art for purposes of determining the patentability of inventions in ways that have complex effects on incentives for information disclosure. Prospective patent applicants typically defer disclosure until after their own application filing dates; otherwise, their own disclosures may become part of the prior art that limits what they may patent.[66] On the other hand, those who wish to defeat potential patent applications of their scientific or commercial rivals may disclose information promptly in the hope of creating more prior art.[67]

Sometimes, however, the goal of defeating potential patents has led to more rather than less restrictive data disclosure policies in public science projects. An example is the international Human HapMap project, which aims to define patterns of common genetic variation across the genome and to make the data freely available in the public domain without intellectual property constraints.[68] The first step was to identify single nucleotide polymorphisms (SNPs) across four human populations with African, Asian, and European ancestry. A further goal is to analyze the information to identify regions in human chromosomes where genetic variants are shared by different individuals, called haplotypes. (Haplotypes are combinations of SNPs that are likely to be inherited together by individuals and are therefore particularly informative to scientists.) The resulting catalog is designed to be a useful resource for scientists who are studying the genetic basis of disease.

Participants in the HapMap project were able to reach an agreement to forebear from filing patent applications on their own raw data. But they were concerned that if they disclosed their data freely, other researchers outside the public project could combine the disclosed data with their own private information and file patent applications on

---

[65] This requirement, referred to as the "Bermuda Rules," arose from an agreement entered into at the International Strategy Meeting on Human Genome Sequencing held in Bermuda in 1996. See David R. Bentley, *Genomic Sequence Information Should Be Released Immediately and Freely in the Public Domain*, 274 SCIENCE 533 (1996).

[66] 35 U.S.C. §§ 102, 103. An inventor's own disclosures will not defeat the novelty of an invention under U.S. law because they do not show prior invention, knowledge or use by another prior to the invention date, 35 U.S.C. §§ 102(a), (g), but they may nonetheless give rise to a "statutory bar" against a patent if the disclosure occurred more than a year before the inventor's filing date. 35 U.S.C. §§ 102(b).

[67] Gideon Parchomovsky, *Publish or Perish*, 98 Mich. L. Rev. 926 (2000); Douglas Lichtman, Kate Kraus & Scott Baker, *Strategic Disclosure in the Patent System*, 53 Vanderbilt L. Rev. 2175 (2000); Rebecca S. Eisenberg, *The Promise and Perils of Strategic Publication to Create Prior Art: A Response to Professor Parchomovsky*, 98 Mich. L. Rev. 2358 (2000). This is a risky strategy, because information that is disclosed in the public domain may also be combined with privately held information to help potential patent applicants to complete their inventions.

[68] The International HapMap Consortium, *The International HapMap Project*, 427 Nature 789-96 (2003)

haplotypes derived from the combined data.[69] In other words, free disclosure of data could lead, paradoxically, to more restrictions on access to haplotypes in the future by helping patent seekers to complete their inventions.

The HapMap project decided to minimize the risk of future patents arising by initially making their data available only to those who agreed to the terms of a click-wrap license agreement modeled after the General Public License developed for open source software (GPL).[70] Under this agreement, those who accessed the HapMap Genotype Database agreed that they would not "take any action (including patenting) that would restrict the access of others to the data within it" or "share the data with anyone who has not read these terms and conditions and agreed to them." In other words, the agreement restricted data-sharing in order to defeat patents. This policy was eventually superseded in December 2004 by an open access policy that puts all data promptly in the public domain.[71]

A similar restricted access policy has also been implemented for the GAIN Program, a public-private partnership run by the Foundation for the National Institutes of Health with the goal of understanding the genetic factors that influence risk for complex diseases.[72] Like the GPL, the GAIN data access policy is designed to defeat future third party intellectual property claims and uses a "viral" strategy of prohibiting further data distribution to anyone who has not agreed to be bound by the same terms. Data are made available at no cost to approved users who agree to the terms of access set forth in the GAIN Data Use Certification document, and who certify that they will not distribute data to third parties who have not agreed to abide by the same terms.[73] By signing the Data Use Certification, those who request access to data acknowledge that GAIN "urges users to avoid making IP claims on the data," while also recognizing "the importance of the later development of IP on downstream discoveries, especially in therapeutics, which will be necessary to support full investment in products that the public needs."

The HapMap and GAIN examples illustrate some of the difficulties involved in adapting the GPL to the release of research data.[74] First, the GPL is structured as a license to intellectual property rights. In the context of open source software, the licensed rights consist of copyright in software, a right that has been recognized by Congress and the courts. Under U.S. law, there is no comparable intellectual property right in data to anchor the HapMap and GAIN licenses. The HapMap license denies this difficulty, requiring those who would access the data to acknowledge, contrary to legal authority,

---

[69] Users of the public data who did not share their own data would have an advantage in identifying haploptypes before those who had access to only the public data.

[70] This agreement, which is no longer in effect, is still posted at < http://www.hapmap.org/cgi-perl/registration>.

[71] http://www.hapmap.org/datareleasepolicy.html.en

[72] http://www.fnih.org/GAIN/GAIN_home.shtml

[73] Updated GAIN Data Access Policy, posted at <http://fnih.org/GAIN/policies.shtml>.

[74] For a general discussion of "open source" approaches in biomedical research, *see* Arti K. Rai, *Open and Collaborative Research: A New Model for Biomedicine*, *in* INTELLECTUAL PROPERTY RIGHTS IN FRONTIER INDUSTRIES: SOFTWARE AND BIOTECH (Robert Hahn, ed., 2005).

that the data are protected by U.S. copyright law.[75] Second, because there is no property right that survives disclosure to those not bound by the license, in order to ensure that third parties do not gain access to the data without agreeing to the terms of the license, the licenses impose strict restrictions on dissemination. Researchers who accessed the HapMap data prior to December 2004 could not release the data to anyone who was not bound by the same license terms. Most notably, they could not include the data in publications based on the data.[76] Third, the GPL is designed to preclude all downstream restrictions on dissemination, an approach that is possible in the area of software, where intellectual property has never been a particularly strong driver of R&D.[77] In contrast, in the biopharmaceutical area, patents – particularly downstream patents on therapeutics – are clearly important. The HapMap license seeks to avoid imperiling downstream patents that might matter for future product development through the use of complex and ambiguous license provisions,[78] while the GAIN license articulates in broad terms its goals and expectations while leaving it to users to figure out how to "avoid premature claims on pre-competitive information, while promoting opportunities to develop IP and file claims on downstream discoveries which will be necessary to support full investment in products that the public needs."[79] Finally, the enforceability of open source licenses remains a somewhat open question. If a public funding agency were to bring a breach of contract action against a license violator, the remedy would be unclear. Perhaps alleged infringers of patents obtained or enforced in violation of the agreement could assert that the patents were invalid or unenforceable for inequitable conduct, but there is no clear

---

[75] *See* HapMap License ¶ 5 ("You acknowledge that the Genotype Database and the data contained in it, to which access is provided under the terms of this License, are protected by law including, but not limited to, copyright laws of the United States. . .").

[76] Data Access Policy for the HapMap Project, http://www.hapmap.org/cgi-perl/registration, [hereinafter HapMap Data Policy] ¶ G ("[while] you are free to publish the results of those analyses [of genotypic information], you may not include in such publications the details of the individual genotypes that the Project has not yet released.").

[77] Although the GPL is designed to preclude all restrictions on dissemination, commercial software developers have deployed other hybrid models that combine open source and proprietary elements. There is a rich literature analyzing these hybrid strategies. See, e.g., Tineke M. Egyedi, (2001), Compatibility Strategies in Licensing, Open Sourcing and Standardisation: The case of Java™, in H. Coenen, M.J. Holler & E. Niskanen (eds) 5th Helsinki Workshop on Standardization and Networks, 13-14 August 2000. http://www.tbm.tudelft.nl/webstaf/tinekee/CompatibilityStrategies.pdf; Franke N., and E.Von Hippel, (2003), Satisfying Heterogeneous User Needs via Innovation Toolkits : The Case of Apache Security Software, *Research Policy* Vol 32:7, 1199-1215; Gabriel R., and W. Joy, (1998), Sun Community Source License Principles,< http://www.sun.com/software/communitysource/SCSLPrinciplesPublic.pdf>; Garud R., and A. Kumaraswamy, (1993), Changing competitive dynamics in network industries : an exploration of Sun Microsystems' open systems strategy, *Strategic Management Journal*, 14, 351-369.

[78] *Compare* HapMap Data Policy ¶ E ("This licensing approach is not intended to block the ability of users to file for intellectual property protection on specific haplotypes for which they have identified associated phenotypes, such as disease susceptibility, drug responsiveness, or other biological utility, as long as public access to, and use of, the data produced by the HapMap Project is preserved.") *with* HapMap License at ¶ 2.b.ii. ("you shall not file any patent applications that contain claims to particular uses of any SNP, genotype or haplotype data obtained from the Genotype Database or any SNP, haplotype or haplotype block based on data obtained from, the Genotype Database, unless such claims do not restrict, or are licensed on such terms that they do not restrict, the ability of others to use at no cost the Genotype Database or the data that it contains for other purposes.").

[79] GAIN Intellectual Property Policy.

authority for such an argument. Such agreements are probably better understood as efforts to define norms of forbearance from enforcement of intellectual property rights within a scientific community than as binding agreements that are themselves enforceable in a court of law.

One can only speculate about the impact of these patent-defeating restrictions on access to biomedical databases. Given the ambiguities in their language and the legal and practical challenges involved in enforcing their restrictions, their primary practical significance may lie in their impact on community norms. Nonetheless, to the extent that the restrictions deter users from accessing the data or from making full use of it, they have a clear downside. Indeed, these patent-defeating restrictions seem to present the same sorts of dynamic losses as the proprietary restrictions on access to the Celera human genome database condemned by the National Research Council Committee on Responsibilities of Authorship in the Life Sciences. To the extent that commercial users feel unable to combine the restricted data with their own proprietary data, potential future discoveries are at least deferred. On the other hand, perhaps this loss is offset by the gains in freedom for future scientists whose work might otherwise be impeded by the future patents that these commercial users would have obtained. But the magnitude of this offsetting benefit is at least called into question by recent survey evidence suggesting that the impact of patents on academic research is generally minimal. If restrictions on access to data are a common obstacle to research, while patents are an uncommon obstacle, a policy that further restricts access to data in order to defeat patents might on balance do more harm than good.

In the rhetoric of patent law decisions, early access to information is so important that it makes is worthwhile for the public to endure "the embarrassment of an exclusive patent."[80] This balance is presumptively reversed in the logic of recent GPL-inspired patent-defeating restrictions on access to databases. Rather than enduring patents to get free access to data, the authors of these restrictive policies have chosen to endure the embarrassment of restrictions on data access in order to defeat patents. The wisdom of this reversal is open to debate.

Conclusion

The ecosystem of public and private research science has grown considerably more complex over the past half-century. The public-private divide in science has blurred as universities have become players in the patent system and as profit-seeking firms have become more involved in upstream research. But even as their research goals and appropriation strategies have sometimes converged, academic scientists have struggled to define their norms and practices so as to distinguish their enterprise from that of their profit-seeking rivals. The resulting interactions offer fertile territory for exploration by the successors of Nelson and Arrow for many years to come.

---

[80] This phrase, attributed to Thomas Jefferson, appears in several patent law opinions of the U.S. Supreme Court. See Bonito Boats v. Thunder Craft Boats, 489 U.S. 141, 148 (1989); Kewanee v. Bicron, 416 U.S. 470, 489 (1974); Graham v. John Deere, 383 U.S. 1, 9 (1966); Marconi Wireless Tel. v. U.S., 320U.S. 1, 60, 61 n.1 (Frankfurter, J., dissenting in part) (1943).