POLICY FORUM: INTELLECTUAL PROPERTY RIGHTS

Database Protection: Is It Broken and Should We Fix It?

Stephen M. Maurer and Suzanne Scotchmer

In 1996, the European Union (EU) issued a directive requiring member states to prohibit unauthorized copying of databases and threatened the United States with retaliation if it failed to enact similar legislation (*1*). Except for opposition from the scientific and engineering communities, the United States probably would have signed a database protection treaty in 1997 and adopted corresponding domestic legislation in 1998 (*2*). A revised bill known as H.R. 354, the Collections of Information Antipiracy Act, is currently pending in Congress.

Collections of Information Antipiracy Act

U.S. courts were historically divided on the question of whether copyright protection extended to databases (*3*). In 1991, the U.S. Supreme Court resolved this issue in *Feist Publications v. Rural Telephone Service Co. Inc.* (*4*). Today, databases cannot be copyrighted unless they demonstrate minimal creativity. Even then, copiers can freely extract facts so long as they rearrange them. H.R. 354 would rewrite *Feist* and bring the United States and the EU closer together by striking a so-called *sui generis* compromise between the current rules and full copyright protection. H.R. 354 would prevent any extraction or use that harms an underlying database's "actual or potential market" (see the box). Protection would last for 15 years. Although H.R. 354 contains exceptions for nonprofit educational, scientific, and research uses, these would not apply where the new use "directly" harmed an underlying database's "actual market."

Scholars have been troubled by the fact that the core concept of a "database" is essentially limitless (*5*). H.R. 354 tries to remedy this defect by limiting protection to information "collected . . . for the purpose of bringing discrete items of information together in one place . . . so that users may access them." However, this standard still literally includes such unlikely "databases" as nonfiction books and movies, both of which "bring discrete items of information together in one place." Congress has relied on judges' common sense to resolve similar ambiguities in the past. Nevertheless, H.R. 354 could lead to sweeping changes throughout intellectual property law.

Today's Databases

Many areas of science and engineering rely on enormous databases as ubiquitous and even indispensable tools. One nuclear science database contains results drawn from more than 160,000 separate references. Some gene-sequencing repositories record more than 2 billion base pairs (*6*).

H.R. 354 AT A GLANCE

What would be prohibited:	Extracting or using "all or a substantial part...of a collection of information gathered, organized, or maintained by another person through the investment of substantial monetary or other resources, so as to cause harm to the actual or potential market of that other person..."
What would be allowed:	• "[E]xtracting or using information for nonprofit educational, scientific, or research purposes," unless the market for the underlying database would be "directly" harmed. • Use or extraction "done for...teaching, research or analysis," unless the result is "likely to serve as a market substitute" for the underlying database. • Reduction or remission of damages against an employee of a nonprofit institution who in the opinion of the judge "had reasonable grounds for believing that his conduct was permissible...."
Duration:	Protection would last for 15 years after data were first offered for sale or placed in commerce.

Listing facts is only the beginning. Some scientific databases reconcile conflicts in the literature, check published calculations, and infer new information by combining reported results. Other databases, particularly in biotechnology, have been rearranged into powerful computing formats that let scientists find increasingly subtle patterns within oceans of data. Finally, many technical databases can be modified and reworked to serve broad new audiences—for example, by turning highly specialized physics data into products that medical doctors, power plant engineers, or advanced sensing device manufacturers can use (*7*). Society will need these activities if it wants to squeeze full value from its investment in science.

Existing Protections

The usual argument for statutory protection sounds simple and compelling. Databases are expensive to make but cheap to copy. For this reason, private and commercial database owners cannot compete with copiers in an open market. If databases cannot earn a fair return under existing law, no rational business would invest in them until Congress changed the rules. Instead, databases flourish (*7*).

This paradox can be explained by remembering that statutes are not the only way to protect databases. Some methods are technological. These include encryption, passwords, and Web sites in which users submit searches without ever seeing the database itself. Other methods depend on existing law. For example, the Copyright Act provides broad protection for some databases (such as electronically searchable versions of *Science*) and limited protection for others. Contract law lets database owners make and enforce promises to keep data confidential.

Finally, many of the most popular and powerful methods depend on the marketplace. If consumers want frequent updates, a would-be copier has little to gain by offering last month's database at a bargain price. Similarly, consumers may think that a particular database is more valuable if it comes with copyrighted search software. In either case, copiers can only compete by making substantial investments of their own. The resulting protection is particularly effective in the sciences, where up-to-date, searchable data sets are at a premium (*7*). Congress could strengthen these methods still further by protecting each update

S. M. Maurer is a practicing attorney at 2632 Hilgard Street, Berkeley, CA 94709, USA. E-mail: maurer@econ.berkeley.edu. S. Scotchmer is in the Department of Economics and the Goldman School of Public Policy, University of California, Berkeley, CA 94720, USA. E-mail: scotch@socrates.berkeley.edu

or correction for 1 to 2 years. Such legislation would be far less restrictive than H.R. 354's proposed 15-year period.

Potential Benefits

Statutory protection should only be enacted if projected benefits outweigh costs. In the case of database protection, three potential benefits need to be considered. First, statutory protection could encourage firms to create new databases that do not exist today. One concrete test is whether existing providers have ever decided against investing in a new database because they thought that someone might copy it. On the basis of a recent report (*7*) and workshop (*8*), such decisions are probably quite rare.

Second, statutory protection could persuade firms to disseminate existing data more widely. Some owners use elaborate secrecy agreements to protect their databases. Such procedures are costly and tend to exclude small users. Other strategies (such as encryption) limit access in ways that some consumers find objectionable. Unfortunately, there is no way of knowing how many database owners would give up secrecy if statutory protection became available. The possibility of patent protection has not put an end to trade secrets.

Third, legislation would reduce waste if providers' efforts to protect themselves through updates, search software, and other enhancements turned out to be exorbitant or led to underinvestment in the databases themselves. Luckily, this problem only arises when fear of copying persuades owners to provide more features than they otherwise would have. At least in the sciences, where the explosion of information has created a desperate need for updates and enhancements, this is unlikely. Providers are likely to go on offering these features whether or not statutory protection is enacted. Although H.R. 354 would probably make copying more difficult and increase the market value of existing databases, proponents of statutory protection should also show that legislation would lead to new or better databases. We now turn to the "costs" side of the ledger.

Inadvertent Disincentives and Pressures to Privatize

Large scientific databases are typically created by combining, extending, and modifying previously published data sets (*7*). H.R. 354 would bar these practices whenever the new products are likely to attract a significant number of users away from their predecessors. Such situations occur frequently. Industrial scientists would typically resolve such conflicts by buying licenses. But how would their counterparts in government and academia respond?

First, they could buy the right to use existing data sets with government grants. In theory, government should be willing to treat this as just another cost of research. In practice, funding agencies are notoriously reluctant to pay for intellectual property licenses. This attitude is unlikely to change (*9*).

Second, scientists could pay for the right to use existing data by selling their own data sets for profit. This could lead to a wave of commercialization. It would also mark a significant departure from current practice, in which most scientific disciplines, including relatively commercialized fields like biotechnology, still rely on a complex mix of commercial and nonprofit databases (*7*). Discouraging individuals and entities that want to provide nonprofit databases seems perverse. Furthermore, fear of giving away commercially valuable data has already injected lawyers into scientist-to-scientist exchanges at universities and national labs. Some U.S. scientists have already begun to complain that the European statutes are making foreign collaborators more reluctant to share data.

Finally, scientists could decide that acquiring all of the rights needed to build a particular database isn't worth the effort. Some biotechnology databases would have to negotiate more than 100 separate contracts. Even large corporations doubt that this is practical (*8*). Obstacles to contracting could make statutory protection counterproductive and even reduce the number of new databases produced (*5, 10*).

Impoverishing Science

The decline of nonprofit databases would force researchers to buy more expensive commercial products. Government sponsors will be reluctant to pay for this. Unless this political difficulty can be overcome, license fees will have to be paid from existing grants and researchers' purchasing power will fall (*9*).

However, statutory protection could also lead to price gouging. Science is particularly vulnerable because many markets are served by sole-source providers. Skeptics argue that potential competition will limit price gouging with or without statutory protection (*11*). Statutory protection could raise the barriers that keep additional competitors from entering the market; if so, fears of price gouging could be well-founded. Little empirical work has been done to determine how many markets are at risk (*11*). However, a closely related industry, scientific journals, has recently been accused of price gouging (*12*).

Congressional Options

It is far from clear that the benefits of statutory protection outweigh the costs. We believe that Congress's best option is to pass no legislation. Failing that, Congress should understand that it is logically impossible to draft a statute that reconciles the monetary incentive structures of commercial life with an academic reward system based on attribution. Probably the best compromise would be to make commercial copying illegal while leaving traditional scientific and research activities alone. This could be done by amending H.R. 354 so that its present exemptions applied whether or not the underlying database's actual market was harmed.

The EU's threats do not change this analysis. It may turn out that American companies can compete in Europe whether or not the EU agrees to protect them from copiers. If not, the United States has ample legal and diplomatic means to challenge the EU, including trade sanctions. Until these options have been tried, Congress should not pass ill-advised legislation.

Database legislation could turn out to be a radical experiment. The U.S. economy has never tried to operate under a statute where most "collections of information" were protected. In the absence of clear evidence that a statute is needed, the traditional bias in favor of keeping data in the public domain ought to be decisive.

References and Notes

1. European Union Council Directive 96/9/EC, O.J. (L77) 20 (1996).
2. W. Gardner and J. Rosenbaum, *Science* **281**, 786 (1998).
3. H. B. Abrams, *The Law of Copyright* (West Publishing, St. Paul, MN, 1998). Courts in traditional U.S. intellectual property centers (New York and California) did not share this view.
4. *Feist Publications v. Rural Telephone Service Co. Inc.*, 499 U.S. 340 (1991).
5. J. H. Reichman and P. Samuelson, *Vanderbilt Law Rev.* **50**, 51 (1997).
6. R. B. Firestone, V. S. Shirley, C. M. Baglin, S. Y. F. Chu, J. Zipkin, *The Table of Isotopes* (Wiley, New York, ed. 8, 1996); National Institutes of Health, *GenBank Overview*, available at www.ncbi.nlm.nih.gov/Web/Genbank/index.html.
7. S. Maurer, "Raw Knowledge: Protecting Technical Databases for Science and Industry," in *Promoting Access to Scientific and Technical Data for the Public Interest: An Assessment of Policy Options,* National Research Council Project CPSM-L-97-05-A, available at www.nas.edu.
8. Workshop on Promoting Access to Scientific and Technical Data for the Public Interest, 14 to 15 January 1999, National Research Council, Washington, DC; proceedings and report available at www.nas.edu.
9. National Research Council, *Bits of Power: Access to Global Scientific Data* (National Academy Press, Washington, DC, 1997).
10. M. A. Heller and R. S. Eisenberg, *Science* **280**, 698 (1998); S. Scotchmer, *J. Econ. Perspect.* **5**, 29 (1991).
11. L. Tyson and E. Sherry, *Statutory Protection for Databases: Economic and Public Policy Issues* (unpublished report, Information Industry Association, 1997), available at www.infoindustry.org.
12. M. J. McCabe, *Academic Journal Pricing and Market Power: A Portfolio Approach* (unpublished manuscript, Georgia Institute of Technology, 1998), available at www.econ.gatech.edu/~mmccabe/journal2.pdf.