

History of Database Protection: Legal Issues of Concern to the Scientific Community

Anne Linn¹
March 3, 2000

SUMMARY

At the instigation of the publishing industry and international information conglomerates, new database protections are being considered in the United States and abroad. In fact, a strong new right for database owners is already being implemented in the 18 countries of the European Economic Area. The EU Database Directive directs member countries to enact laws preventing unauthorized use of more than insubstantial portions of a database for 15 years after the database was produced. (In some countries, scientists and teachers are permitted to use substantial parts of databases as long as their activities are not commercial in nature.) In response to these factors, several bills have been introduced in Congress and an international treaty has been put forward in the United Nations' World Intellectual Property Organization. Because of strong coordinated opposition from scientific and library organizations, none of these efforts have succeeded to date and Europe is the only region in the world with these new protections.

New database protections, however, are supported by both houses of Congress and the administration, so legislation will likely pass this year in the United States. (Two bills are currently pending in the House.) If it does, the World Intellectual Property Organization will likely resume work on a global treaty. In the absence of U.S. opposition, such a treaty could pass, resulting in new database protection in 171 countries. The U.S. legislation currently being considered is somewhat different in design from the EU Database Directive, but the impact on science and education would be similar. Unless substantial modifications are made, it could lead to a more restricted environment for data collection, exchange, and use. In particular, enactment of U.S. database legislation under discussion could have the following impacts:

¹ This background paper expresses the views of the author and not necessarily those of the National Academies or the National Research Council. Please send comments to alinn@nas.edu.

- reduce the amount of data that can be obtained, particularly from the private sector or public-private partnerships, an increasingly important source of data;
- increase the cost of obtaining data, particularly from database owners with a monopoly on the data;
- restrict access to data for at least 15 years from the time the database was created;
- discourage the transformation of existing databases into new ones, creating artificial gaps in data availability;
- prevent the use of data for purposes other than which it was collected, minimizing the scientific and societal value of the original data; and
- increase restrictions on the use of compilations of all kinds, including works of authorship (e.g., collection of articles) not normally considered to be databases.

Further restrictions on the acquisition and use of data are likely to be placed on researchers by risk-adverse universities and government agencies seeking to avoid the possibility of costly litigation. The net result is that a legal culture would be created which encourages commercial exploitation at the expense of the public domain.

DATABASE PROTECTION IN THE UNITED STATES AND EUROPE

What are the Drivers?

Databases are protected against piracy through a combination of legal and technical means—primarily copyright and contract, but also patent, trademark, trade secrets, and encryption. This legal and technical environment, however, has changed significantly in recent years because of the following factors:

- digital environment—individuals can now copy and distribute publications and large amounts of data at little cost or effort;

- U.S. Supreme Court *Feist* decision and similar decisions in European high courts—restated copyright law principle denying protection to databases produced by *sweat of the brow* (i.e., databases created with large amounts of money, effort, or labor) but without creativity; and
- European Database Directive—provides 15 years of protection for the contents of the database and each significant update, and permits database owners to prevent the use of substantial parts of the database. The directive also has a reciprocity clause which states that only countries which offer similar protections to EU nationals will receive this new level of protection within the European Economic Area.

Notwithstanding the *Feist* decision, most databases are protected by copyright, which protects the creative elements of a database—the selection, coordination, and arrangement of the information—although not the facts themselves. For example, the yellow pages are protected by copyright because the organization of information, use of boxes, colors, etc., required thought and creativity. On the other hand, the white pages are a simple alphabetical listing, which is *not* protected by copyright. Most databases used by scientists either fall under copyright law or are in the public domain and available to all. (By law, the U.S. federal government cannot copyright databases, although private vendors disseminating government information can.) Scientists can generally use copyrighted material because of a fair use exception in the United States or equivalent exemptions in Europe (see Box 1).

Box 1. Fair Use Exceptions in U.S. Copyright Law

Fair use is a bedrock principle that reconciles the Copyright Act’s grant of exclusive rights to authors and the First Amendment’s constitutional guarantee of free speech. Under copyright, certain public purposes including “criticism, comment, news reporting, teaching, scholarship, or research” are permitted.

U.S. courts consider four factors for determining whether the fair use exception is allowable:

- the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
- the nature of the copyrighted work;
- the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- the effect of the use on the potential market for or value of the copyrighted work (most heavily weighted factor).

Decades of court interpretations define what is meant by fair use today. A fair use exception is most likely to be granted under the following conditions:

- the use is for non-commercial purposes;
- the original work was inexpensive to produce and/or distribute;
- a relatively small amount of the original work is used;
- the portion used is transformed, not merely copied; and
- the economic impact is insignificant.

It is important to note that the level of protection offered by copyright is thin compared with the new additional protection offered by the EU Database Directive or proposed U.S. legislation. Even without such additional protection, however, other legal means (e.g., contract) and technical devices (e.g., encryption) can be used by database producers to maintain control over unauthorized use of a database. A contract is a two-party agreement, the terms of which are specified by the individuals involved. It can be used to prevent unauthorized uses of a database by the parties to the agreement. This form of protection has some limitations, including (1) a high administrative burden of negotiating terms with each user and provider of data, particularly for databases compiled from several sources, and (2) they cannot prevent unauthorized downstream uses of the database because they are only binding on the parties to the agreement. Downstream uses of databases can be controlled through encryption, although such measures can be expensive or cumbersome to implement.

In general, the existing legal regime of copyright plus contract protects databases produced and distributed within the United States. U.S. databases distributed in Europe will also receive copyright and contract protection, but not the stronger legal protections of the EU Database Directive. Because of the reciprocity features of the EU Database Directive, unless similar legislation is enacted in the United States, U.S. databases could theoretically be at a competitive disadvantage in Europe, where they may be susceptible to unauthorized uses.

EU Database Directive

The EU Database Directive was created to harmonize the intellectual property laws regarding databases of the 18 countries of the European Economic Area (EEA) by supplementing copyright

to protect databases produced by sweat of the brow. The Directive was passed in March 1996 and member nations were responsible for implementing it by January 1998, although only nine countries have implemented it so far (see Table 1 below). The Directive creates a new kind of intellectual property protection (a *sui generis* right, which means of its own kind) for databases produced in the EEA. Under the Directive, database producers can prohibit use of more than an insubstantial part of the database. The term of protection is 15 years, but each time the database is updated significantly, the entire database (not just the updated parts) receives another 15 years of protection. Consequently, active databases apparently can be protected in perpetuity.

Member countries are permitted to designate exceptions and limitations in their implementing legislation, as long as the exceptions do not conflict with the normal exploitation of the database. (Anyone may use insubstantial parts of the database for any purpose.) Most countries which have implemented the directive have granted exceptions for science and education as long as these activities do not serve a commercial purpose (Table 1). (France does not permit any exceptions.) This is a narrower exception than that granted in copyright. In addition, many EU countries have freedom of information acts, which provide for access to government data, but it is not clear whether they can be overridden by the Database Directive. Moreover, freedom of information acts do not include data collected or disseminated by state-owned companies operating under market conditions without a public service obligation. Finally, basic principles of European law may in some cases constrain the ability of database makers from exercising monopoly control over information protected by the Database Directive.

The European Commission is supposed to review the impact of the Database Directive in 2001. As input, member nations will report on whether the *sui generis* right has decreased competition. If so, non-voluntary licensing agreements may be imposed on database producers to increase user access.

Table 1. Implementation of EU Database Directive

| EEA Country | Date Implemented | Exceptions for Science and Education |
|--------------------|-------------------------|---|
| <i>Implemented</i> | | |
| Austria | January 1, 1998 | Substantial use permitted for scientific purposes without the intention of commercial exploitation; data can be published |

| | | |
|------------------------|-------------------|--|
| | | if source is acknowledged |
| Belgium | September 1, 1998 | Not available in English |
| Denmark | July 1, 1998 | Not available in English |
| Finland | April 4, 1998 | Not available in English |
| France | June 16, 1998 | No exceptions for science or education |
| Germany | January 1, 1998 | Substantial use permitted for private purposes, teaching in non-commercial institutions, and scientific purposes to the extent that copying is necessary and does not serve commercial purposes; data can be published if source is acknowledged |
| Spain | April 1, 1998 | Not available in English |
| Sweden | January 1, 1998 | Not available in English |
| United Kingdom | January 1, 1998 | Substantial use permitted for purposes of illustration in teaching and research, as long as the purpose is non-commercial and the source is indicated [data publication is not explicitly permitted]; public records are also exempt |
| <i>Not implemented</i> | (as of June 1999) | |
| Greece | | |
| Ireland | | |
| Italy | | |
| Luxembourg | | |
| The Netherlands | | |
| Portugal | | |
| <i>Unknown</i> | | |
| Iceland | | |
| Norway | | |
| Liechtenstein | | |

History of U.S. Database Legislation

In recent years, four database bills have been introduced in Congress, although none have become law. The key provisions of the failed bills are summarized in Box 2. The first U.S. database bill (HR 3531) was introduced in the House in 1996 and was modeled after the EU Database Directive, except that the term of protection was 25 years, instead of 15 years. Like the EU Database Directive, there were no exceptions for fair use or for government data. The bill also imposed potentially severe criminal penalties. These provisions alarmed the scientific and library communities, which sent letters to Representative Moorhead expressing their concerns

and asking for a period of public debate. Supporters of the legislation were surprised by such strong opposition and the bill was not brought to a vote.

A year later, the second database bill (HR 2652) was introduced in the House. HR 2652 was slightly more science-friendly than the previous bill, but users were still forbidden to use more than insubstantial parts of a database, and they would be punished if their actions resulted in economic harm to the database producer. Given the high cost of some data sets (e.g., a single synthetic aperture radar scene costs \$1,600), economic harm would be easy to prove. HR 2652 also permitted exceptions for government data and for scientific uses, but the exceptions did not apply if a *potential* market might be harmed or if the data were collected by public-private partnerships (e.g., SeaWiFS), an increasingly important method of data collection. A hypothetical example of the impact of HR 2652 on genetic research is given in Box 3.

Box 2. Key Provisions of Failed U.S. Database Bills

1996 HR 3531 (Moorhead)—*sui generis* approach (creates a new property right)

- 25 year term of protection
- criminal penalties
- no fair use exceptions
- no exception for government data

1997 HR 2652 (Coble)—“so called” misappropriation approach (harm triggers liability)

- 15 year term of protection
- no criminal penalties to non-profits
- exception for non-profit science unless harm to *potential* markets
- exception for government data unless overridden by contract or collected by public-private partnership

Hearings on HR 2652 were held, but the invited scientists failed to make their case and the bill passed the House unopposed. The bill was subsequently folded into a House copyright bill, which also passed, then moved into House-Senate conference. Up to this point, the strategy of database opponents—a loose coalition of scientific groups, libraries, telecommunication companies, Internet service providers, and value-added database producers—was to question the need for additional database protections, given the absence of documented cases of database piracy and the likely harm to science and education. With the arrival of the copyright/database

bill in conference committee, opponents began drafting alternate language for a database bill. A compromise could not be reached and the database provisions were ultimately thrown out. (The copyright provisions passed subsequently and are now law.)

Box 3. Hypothetical Example Illustrating Pitfalls of HR 2652

Advanced Genetic Data (AGD) has compiled data on variation in human DNA sequences and sells access to these data to pharmaceutical firms and other biotechnical customers. The company made a considerable investment to compile the database from their research and from publicly available databases.

Dr. Susan Jones is a molecular geneticist funded by NIH. She has developed a software application that detects whether DNA samples contain members of a library of biologically significant target sequences. The sequence library is stored in a database that is a component of Dr. Jones's application. Dr. Jones compiled her library from various sources, including sequences purchased from the AGD database. After the publication of her work, she shared her application, including the sequence library, with colleagues working on similar problems.

At about the same time, AGD tripled the price of accessing the data set. AGD also filed suit against Dr. Jones, stating that by sharing AGD's sequence with colleagues, she has harmed their market for the data themselves and the software application embodying the data that AGD had planned to develop.

SOURCE: Gardner, W. and J. Rosenbaum, 1998, Database Protection and Access to Information. *Science*, vol. 281, p. 786-787)

Potential Consequences to Science

By turning data into a commodity, the database protections in force or currently being considered will likely exacerbate problems U.S. researchers are having with existing commercialization policies in other countries. Organizations with commercialization policies rely on contract law to restrict the use of data to approved individuals and/or for specific purposes. Such contracts commonly prohibit normal scientific practices, such as sharing the data with colleagues, publishing the data in scientific journals, or using the data to address several different scientific problems. Contracts can also be written to override fair use exceptions in most cases.

With the passage of the EU Database Directive, European commercial database producers and privatized government agencies have a new tool for restricting data use. However, the directive

does not permit the exceptions for science and education to be overridden by contract. In this sense, the EU Database Directive is more science friendly than U.S. database bills considered thus far, none of which have imposed limits on contract. Thus, any fair use exception in U.S. database legislation could be overridden by means of contract law.

If database legislation passes in the United States, it will likely have exceptions for government data. Consequently, as long as scientists and teachers obtain their data directly from the U.S. government, the proposed database legislation may have little impact on their activities. However, the role of the U.S. government in collecting and disseminating information is changing in the following ways:

- the number of public-private partnerships is growing;
- federal agencies are going out of the data collection business and are increasingly willing to buy data for scientific purposes from commercial vendors; and
- the private sector is becoming increasingly involved in disseminating government data.

These changes are partly a result of declining budgets, which force agencies to look for partners to share costs, and partly a result of new legislation and regulations aimed at reducing competition with the private sector. For example, the Commercial Space Act of 1997 encourages NASA to purchase data collection and dissemination services from the private sector. NASA has already taken advantage of a commercial partner when it teamed with Orbital Sciences Corporation to launch SeaWiFS, an ocean color sensor of interest to the fishing and shipping industries, as well as to oceanographers and global change researchers. Similarly, NOAA has announced that it will no longer allow the data collection systems on its geostationary and polar-orbiting satellites to be used where there are commercial space-based services available that meet the user's requirements.

The increasing involvement of the private sector in scientific data collection and dissemination has two ramifications for science: (1) the resulting data are eligible for copyright and database protections not available to government data sets, and (2) a market for scientific data is

developed. The first is important because less data could enter the public domain, thereby increasing the cost of obtaining data and/or restrictions on its use. The second is important because the argument is circular: where scientific data is concerned, researchers form the commercial market and are therefore ineligible for a fair use exception. (Scientists working at commercial institutions would not be eligible for a fair use exception in any case, even if their research is not directed toward the development of a commercially competitive product.) Thus, researchers freely sharing data and applications could reduce the profits of a data vendor and draw a lawsuit (see Box 3). Even the threat of such lawsuits could undermine the principles of sharing data for the benefit of the community and seeking rewards from publication and attribution. In the long run, selling data to scientists may not prove to be a viable commercial strategy, but as Landsat showed, such commercialization experiments can cause significant setbacks to science.

In addition to these long-term changes to the public domain, passage of the proposed database legislation could have an immediate impact on scientific practice, particularly for basic research with commercial applications. Table 2 compares scientific practices that are permitted under copyright and the database protections of HR 2652. If database legislation passes, scientists and other users would have to conform to both copyright (which protects the creative elements) and database protection (which protects the facts themselves) provisions.

Table 2. Scientific Practices Under Copyright and Database Protection*

| Practices Permitted Under Copyright | Practices Permitted Under HR 2652 |
|---|--|
| <ul style="list-style-type: none"> • use <i>all</i> of the factual data in a database, regardless of the amount or age of data being used, as long as the creative elements (i.e., selection, arrangement) are not directly or indirectly reproduced | <ul style="list-style-type: none"> • use <i>insubstantial</i> amounts of factual data in a database; • use <i>all</i> of the factual data in a database as long as the data are more than 15 years old; or • use <i>all</i> of the factual data in a database as long as all the following conditions are met: (1) the purpose is justified for teaching or research, (2) the individual is at a nonprofit institution, and (3) the action does not harm the market |
| <ul style="list-style-type: none"> • use the creative elements of a database for public purposes such as teaching, scholarship, or research, subject to the fair use doctrine | <ul style="list-style-type: none"> • same as above—database laws don’t distinguish between factual data and creative elements |
| <ul style="list-style-type: none"> • recreating an entire database is prohibited (if | <ul style="list-style-type: none"> • recreate an existing database using data from |

it can't be done without reproducing the creative elements), even if original sources were used

the original sources

- | | |
|--|--|
| <ul style="list-style-type: none">• combine the factual data with other data into a new database without permission or additional payment to the originators• purchase a book or article, then lend it to a colleague• borrow a book or article from a library, use it for virtually any purpose, and make a copy of it for scholarly purposes | <ul style="list-style-type: none">• combine the factual data with other data into a new database as long as permission is obtained and/or payment is made to the originators• purchase access to a database and lend the data to a colleague as long as doing so did not result in potential lost sales• borrow a database from a library as long as it is used for scholarly purposes |
|--|--|

*Note: these practices may also be subject to contractual provisions.

Current Status of Database Legislation in the United States

Database producers, Congress, and the White House have all agreed that additional database protections are needed. Thus, database legislation may be inevitable, and may even be passed during this session of Congress. The White House has identified five principles for database legislation:

- the language should be simple, minimalist, and clear;
- there must be exceptions for government data;
- prohibited activities should be clearly defined to avoid unintended consequences;
- fair use exceptions similar to copyright should be included; and
- U.S. databases should receive the same protections in other countries as databases produced in those countries (i.e., satisfy the reciprocity clause of the EU Database Directive).

Two database bills—HR 354 and HR 1858—are currently pending in the House of Representatives. Each takes a different approach to prevent unfair competition in the form of parasitic copying of databases (Box 4), and would have very different consequences for science and education. HR 354 was introduced in the House Judiciary Committee by Coble in January 1999. It is oriented toward database *producers* and prohibits uses which could harm the primary or related market of the database. On the other hand, HR 1858, which was introduced in the

House Commerce Committee by Bliley in May 1999, is more oriented toward database *users*. HR 1858 allows all uses of databases, except commercial uses meant to compete directly with the original database.

HR 1858 has a broader range of exceptions than HR 354. For example, both bills exclude government data from protection, but HR 1858 also excludes individual ideas, facts, principles, preexisting databases, and works of authorship. Both bills contain fair use exceptions, but HR 354 permits only non-profit uses, and only if they do not result in market harm. HR 1858 permits all scientific and educational uses, including those in the private sector, as long as the database is not used for purposes of direct commercial competition. In addition, systematic or repeated use of a database is permitted under HR 1858, but not under HR 354. For these reasons, HR 1858 is supported by scientific organizations, libraries, value-added database producers, Internet service providers, and telecommunications companies. However, the bill may not satisfy the reciprocity clause of the EU Database Directive.

Both bills have been marked up and are awaiting action in the House. However, HR 354 has more than three times as many supporters as HR 1858 and will likely be the first (if not only) bill to be voted on in the House. To date, there is no corresponding bill in the Senate.

The following advice was voiced by Judge Edward Damich, a former Hatch staffer who worked on database legislation:

The scientific community has done an impressive job of getting organized over the past five years. Their main strengths are (1) no one in Congress wants to be against science and education, and (2) there is widespread recognition that science underpins the economy. On the other hand, scientists tend to be uncompromising and risk being excluded from debate. If the science community wants to play a role in the database issue, it should (1) make the case for science, (2) know what compromises it can live with (and be ready to compromise); and (3) come to the negotiating table with specific language for the bill.

Scientists have begun to heed Damich's advice by helping to draft alternate language for current and past database bills and supporting HR 1858.

Box 4. Key Provisions of Current U.S. Database Bills

HR 354 (Coble)—*so called* misappropriation approach

- broad prohibition of database uses, with narrow exceptions
- 15 year term of protection, with no extension for later updates
- no criminal penalties to non-profits, and reduced or eliminated monetary damages
- exception for non-profit science unless *material* harm to *primary* markets
- exception for government data unless overridden by contract

HR 1858 (Bliley)—*targeted antipiracy* approach

- most database uses are permitted, except those meant to compete commercially
- no term of protection
- no criminal penalties to *anyone*
- exception for science unless the purpose of use is direct commercial competition with the database producer or avoiding payment of reasonable fees
- exception for government data unless overridden by contract
- limitations on database monopolies

GLOBAL DATABASE ISSUES

The World Intellectual Property Organization (WIPO) has been considering database protection since 1996. WIPO is a specialized agency of the United Nations and is responsible for the promotion and protection of intellectual property throughout the world through cooperation and treaties among its 171 member nations. The U.S. Patent and Trademark Office (Department of Commerce) heads the U.S. delegation to WIPO.

Database action in WIPO began in December 1996, when delegations from the European Union and the United States introduced a treaty modeled after the EU Database Directive. A bill with similar provisions (HR 3531) was simultaneously introduced in the U.S. House of Representatives to help ensure U.S. support and passage of the WIPO treaty. As noted above, however, strong opposition from the scientific and library communities led to the withdrawal of HR 3531, and the U.S. Delegation to WIPO was instructed to oppose its own treaty. (One of the most effective letters in changing the U.S. position came from the presidents of the National Academies. They described the proposed bill as having a “deleterious long-term impact on our

nation's research capabilities" by making it difficult for scientists to reuse and combine data for publication or research.)

Since that time, WIPO has sponsored a number of information meetings to gather input from a broader range of stakeholders. Notable among the nongovernmental organizations permitted to attend the information meetings and submit position papers are the World Meteorological Organization (WMO) and the International Council for Science (ICSU). ICSU created an international committee (ICSU/CODATA Committee on Data and Information) to speak on its behalf at these meetings. Both the ICSU committee and the WMO secretariat have written papers opposing the provisions of the proposed treaty and describing the importance of full and open exchange to science and education. The ICSU papers have also defined scientific principles that should be upheld in any database treaty and provided examples of a wide range of research activities that could be adversely affected by such a treaty. The information meetings will continue, although the timetable is unclear. The schedule is likely to be accelerated by passage of a database bill in the United States. U.S. legislation and the EU Database Directive will probably be used as the starting point for a global treaty.

Meanwhile, the ICSU/CODATA Committee on Data and Information is seeking to establish a dialog on the database issue among European scientists, few of whom have ever heard of the EU Database Directive or the proposed WIPO treaty. Participation by scientists in the process is important for determining the impact of databases leaving the public domain as a result of the directive and other commercialization policies. This information would also be valuable input to an eventual WIPO treaty or the 2001 review of the EU Database Directive. Thus far, however, efforts to engage European scientists on this issue have failed.