R A W K N O W L E D G E:

P R O T E C T I N G T E C H N I C A L D A T A B A S E S
F O R
S C I E N C E & I N D U S T R Y

A Report Prepared for the National Research Council's
Workshop on Promoting Access to Scientific and
Technical Data for the Public Interest

January 14 - 15, 1999

Stephen M. Maurer, Attorney-at-Law
2632 Hilgard St., Berkeley, California 94709

Table of Contents

Executive Summary*

As usually defined, "databases" include numerical data, text, images, or any other "organized collection of information." Because enormous numbers of products fit this description, it is sometimes hard to think about such apparently straightforward questions as "Is existing legal protection adequate?" or "Could it be improved?" This report tries to make matters more concrete by examining existing databases and how they are produced. The results are then used as a benchmark to evaluate potential legislation. Special attention is paid to features and problems that set scientific/technology databases apart from other products.

The world of scientific and technology databases is already extremely rich and well-developed. Since the US government has never enacted database legislation, this presents a paradox: If existing databases can be freely copied, why do firms continue to invest in them? The answer is that database providers have devised a bewildering number of unofficial ("self-help") methods for protecting their investments. These include but are not limited to (i) bilateral agreements with users, (ii) "shrinkwrap" or "clickwrap" language, (iii) bundling with copyrighted materials, (iv) continual updating and improvement that leaves would-be copiers "out of date," (v) search-only Web sites where the underlying database cannot be downloaded, and (vi) passwords and encryption. *The fact that rich and diverse databases exist in today's world shows that such protection can be extremely robust.* At the same time, self-help strategies may cause undesirable distortions in the economy, particularly when they discourage database suppliers from sharing products with a wider audience. Even more insidiously, lack of statutory protection may mean that some databases are never created in the first place.

Scientific and technology databases present unique needs and problems. These include:

- The need to assure private firms that they can profitably invest in "commercializing" and extending government databases for use by a broader audience;
- The need to keep database prices within the reach of academic users, who have traditionally driven most advances in basic knowledge;
- The scientific community's need for "value added" or "edited" databases that not only collect but also update, cross-check, comment on, and try to reconcile reported results;
- The fact that virtually all scientific databases have historically been created by

* The author wishes to thank the National Research Council for commissioning this study and to acknowledge helpful conversations with Suzanne Scotchmer, Jeannette Balko, D. Ben Borson, Jack Brown, Richard Firestone, Richard Gilbert, Karl Kenna, Elizabeth Powers, Jerry Reichman, Kenneth Rosenblatt, Pamela Samuelson, John Stattler, Tom Slezak, Paul Uhlir, and Joel White. The author is solely responsible for all opinions, errors, and omissions contained herein.

combining and extending earlier data sets; and

- The scientific community's need for full and unrestricted access to data, which inevitably conflicts with "self-help" strategies based on secrecy or partial disclosure.

The modern history of database reform begins with the US Supreme Court's 1991 decision in *Feist Publications, Inc. v. Rural Telephone Service Co.*, which restricted "sweat of the brow" protection under copyright in the United States. This was followed by the European Union's 1996 Directive on Databases, which required member countries to *expand* their statutory protection of databases. The Directive also contained a controversial threat that citizens of countries (including the United States) that did not adopt EU-style statutes would not be protected by the new laws when they took effect. Because of the EU Directive, the US Congress introduced European-style legislation in 1996 and again in 1997-98. Scholars have also suggested alternatives to the European model.

Existing reform proposals can be broadly summarized as (i) *de minimis* changes to existing law, (ii) "unfair competition" schemes that would examine the need for protection on a case-by-case basis, and (iii) so-called *sui generis* protection that would give database owners strong property rights modeled on the EU Directive. The principal difficulty has been to reconcile these proposals with the "public domain" principle that "mere facts" cannot be protected. Although this is an old problem, courts were frequently able to avoid it in the past because copyright and patent law only protected a small fraction of all possible commercial knowledge. Comprehensive database protection would turn the situation on its head by making virtually all facts protectable as "organized collections of information."

In the final analysis, the policy debate for and against database protection cannot be settled by purely legal considerations. Instead, the underlying question is largely empirical. If free ridership turns out to be a problem for all databases, then some sort of additional protection should be enacted. But if free ridership is only "sometimes" or "never" a problem, reform should be much more cautious. The fact that such questions have so far received relatively little attention makes the Committee's work especially timely and represents a valuable opportunity to advance debate in this area.

Part I: Today's Databases

The concept of a "database" is usually defined quite broadly: For example, one typical formulation describes a database as **"**any organized collection of information"[1] even though the same phrase could just as easily describe intellectual property in general. The problem is that such definitions are too broad to provide a concrete sense of which databases actually exist in today's economy or why they should be protected.

Part I of this report tries to make the concept of a "database" more concrete through examples, anecdotes, and case studies. By way of background, Examples l through 3 describe some non-technical databases that are available on CD-ROM, over the Internet, and in print. Examples 4 through 8 continue the discussion by describing an assortment of databases drawn from the physical sciences, biotechnology, and engineering. The final section ends by collecting and commenting on various "lessons" learned from these Examples. The lessons provide a benchmark for evaluating proposed reforms later in this report.

A. Some Commercial Databases

Example 1: A Sampler of CD-ROMs. As of January 1995, the authoritative *Gale Directory of Databases* listed 9,385 electronic databases for sale by commercial vendors. The list

was further subdivided by format, including "on-line" and "CD-ROM. " Table 1 analyzes a sample of 100 databases randomly selected from the catalog's CD-ROM listings[2]:

A CD-ROM Sampler

Vendor/Type	**Numerical Data & Directories**	**Software**	**Bibliography**	**Text, Image & Multimedia**
Government Provider	6	0	1	0
Commercial Provider of Public Domain Data	0	1	6	8
Commercial Provider of Public Domain Data Enhanced With Proprietary Software Or Other Features.	7	0	0	0
Commercial Provider of Original Data	9	3.0	21	38.0

Table 1

The fact that such a rich and diverse selection of databases has evolved *without* statutory protection is striking. At the same time, Table 1 illustrates the fact that database suppliers use a variety of non-statutory strategies to protect their products. These include:

- **Copyrighted Content.** One of the most surprising aspects of Table 1 is that most products continued to follow "traditional" print-based models. For example, nearly half of the sample (46%) consisted of **Text, Image & Multimedia** -- predominantly electronic versions of books, journals, and newspapers. Virtually all of these materials are individually protected by copyright whether or not they are included in a database.

- **"Free" Counterparts.** Another way to look at **Text, Image & Multimedia** is that the cost of producing them electronically tends to be small *once print-based counterparts already exist.* This makes electronic databases extremely tempting to would-be providers.

- **Updating.** Seventy-three percent of the products listed in the sample were regularly updated on a quarterly or annual basis. This practice makes it extremely difficult for would-be copiers to sell a "current" product.

- **Enhancements.** Many CD-ROM databases were packaged with advanced (and presumably copyrighted) search software. For example, many of the **Numerical Data and Directory** products combined public domain data with advanced software for making customer lists, address labels, or performing searches. Since the copyright laws protect software, the presence of such enhancements forces would-be copiers to choose between selling a visibly inferior product and making investments of their own.

- **Reprints.** Finally, some large providers were able to sell "reprints" of government databases despite the fact that this information is freely copyable. The tactic appears to work because large providers have advantages of scale when it comes to finding widely scattered consumers in a "thin" market. By contrast, would-be copiers are too small to locate these same consumers for themselves.[*]

It is probably significant that the *Gale Directory* showed no obvious difference between CD-ROM products and those available "on-line." As explained below, the Web offers several distinct technical advantages for "self-help" security. The fact that providers chose to forego these advantages shows that security can be accomplished in various ways.

[*] The best example of this is a company called "Silver Platter." Silver Platter's nuclear databases are discussed in my interview with Richard Firestone (Exhibit 3).

Example 2: An Internet Sampler. Table 2 summarizes 100 Web sites obtained by searching for the word "databases" on the "Infoseek" search engine:

An Internet Sampler

Provider Type	**Bulletin Board in Which Individual Needs are Posted a Single Place (*e.g.* job listings)**	**Compilations of Two or More Public Domain Databases**	**Original Data**	**Directory & Network Data Which Identify Community Members to Each Other and/or the Public**
Enthusiast	0	6	0	0
Government & Education	2	40	3	4
Commercial Provider/Access Provided Without Charge	6	0	7	3
Commercial Provider/Portions of Data Restricted to Users Who Have Purchased Passwords.	0	0	8	0
Commercial Provider/Search Only	0	0	18	3

Table 2

Table 2 significantly expands the list of self-help strategies found Table 1. These include:

- **Passwords and Two-Tier Access.** Perhaps the most traditional way to protect databases is to use passwords. Many Web sites provide "free samples" of password-protected data.

- **Search Only Web Sites.** The most common form of "self-help" found in the sample was for users to submit requests to the vendor, who would then perform searches on their behalf. Like passwords, this provides essentially complete protection against piracy.[*]

- **Clearinghouses.** Some databases earn income by selling listings instead of

[*] Tom Slezak, a computer scientist at Lawrence Livermore National Laboratory, confirmed that these methods conferred "reasonable and prudent security" when I interviewed him on November 20, 1998. A memorandum summarizing Slezak's comments can be found at Exhibit 6 to this report.

charging user fees. The classic example is a job agency, in which employers pay for ads which are then distributed to the public without charge.

- **Product Ties and "Come Ons."** Many Web databases are offered as a "free" inducement to purchase related products. In such cases, the producer provides data without charge in order to promote his "core business."[*]

Finally, Table 2 presents a stark reminder that not all database providers *want* protection. This is trivially true for the enthusiast, government, and education providers whose missions are heavily slanted toward dissemination. A more subtle point is that the phenomenon also exists in the commercial field, where databases are frequently used as "market makers" to bring buyers and sellers together. It is an open question whether or not such players *welcome* copying (particularly when they receive attribution) as a way of reducing their own publication costs and/or reaching even larger audiences.[**]

Example 3: Other Types of Databases

Example 3(a): Dataquest. Even though all of the providers found in Table 2 offered more or less standardized products, this is not the only business model available on the Web. For example, a consulting group known as "**Dataquest**" offers two types of proprietary information: (i) a library of 25,000 confidential reports that can be searched and downloaded over the Web at a cost of between $100 and $ 5,000 per item, and (ii) custom research at a negotiated price. All of these products are subject to elaborate contractual safeguards governing each side's use and disclosure of the reports. Dataquest also sells "alert" services that notify users of developments in pre-defined areas of interest.[3]

Example 3(b): Info-Trac. One of the largest (and most useful) databases found in the course of preparing this report was a citation index called Info-Trac. Info-Trac is available both on-line and as a CD-ROM. Although Info-Trac is available to users (*e.g.* libraries) free or at nominal cost, it charges a substantial fee for copying hard-to-find articles.[4] This is yet another example of using an essentially "free" database to market the seller's principal product.

Example 3(c) Paper-Based Databases. The fact that *Feist* involved *telephone books* shows that paper-based databases are still important. Virtually all of the text and bibliographic products listed in Table 1 have print-based counterparts.

---

[*] Perhaps the best example of this in the sample was an "online" video store that allowed users to search a massive database of over 125,000 movies, many of which were not even available commercially.

[**] *See also,* J.H. Reichman & Pamela Samuelson, "Intellectual Property Rights in Data?," *Vanderbilt Law Review* Vol. 50, p. 51 (January 1997) at p. 67 ("To the extent that government generated or university generated data remain noncommercialized, their vulnerability to technically refined means of [copying] may be of relatively little importance. Presumably, the originators want the broadest possible distribution of their datasets.")

B. Some Scientific Databases

Example 4: Some Electronic Databases "Samplers." Most of the examples listed below describe the creation, evolution, and/or capabilities of individual databases. The present section tries to set the stage by presenting broader, more impressionistic "samplers" of scientific and engineering databases offered over the Web or in libraries. Because the samplers contain considerable overlap, they are discussed together at the end of this section.

Example 4(a): Physics. Table 3 extends the previous discussion to the sciences by summarizing on-line and CD-ROM databases offered (i) by the University of California's Berkeley Physics library, and (ii) by the results of a "Yahoo Physics" search engine request for the word "database."*

A Physics Sampler

Resource	Number	Comment
On-Line Versions of Print Journals	55	Includes physics-related journals published by American Physical Society, Reed-Elsevier, American Chemical Society, and American Astronomical Society.
Electronic Preprint Servers	8	Includes servers maintained by government laboratories and professional societies. The American Institute of Physics also offers its own e-journal.
Electronic Abstracting and Indexing Databases	2	Consists of (i) INSPEC database of 4,000 journals plus selected conferences, reports, dissertations, and books, and (ii) Web of Science database of 3,300 scientific and technical journals.
Other Electronic Resources	12	Includes large atomic, particle, and thermodynamic databases prepared by national labs and universities.

Table 3

* The UC Berkeley and Yahoo Web Pages used to compile Tables 3 and 4 can be found at Exhibit 2. Interested readers may want to acquire a "feel" for existing databases by skimming through these listings.

Example 4(b): Engineering (Library Resources). Because of their greater volume, the corresponding UC Berkeley and Yahoo resources for engineering are listed separately. The library resources are as follows:

An Engineering Sampler
(U.C. Berkeley Resources)

Resource	Number	Comment
On-Line Versions of Print Journals	60	Includes journals published by professional societies (ACM, ACS, IEEE, Society of Industrial and Applied Mathematics) and private publishers (Academic Press, Elsevier, Springer, Wiley).
Electronic Abstracting and Indexing Databases	16	Includes private, DOE, EPA, and National Technical Information Service publications
Technical Report Databases (Includes both indexed and full text)	10	Includes government sites and the Yahoo Physics search engine.

Table 4

Example 4(c): Internet Databases for Engineering. Finally, Table 5 summarizes the seventy-one relevant "hits" generated by polling the Yahoo Engineering search engine for the word "database":

A Second Engineering Sampler
(Internet Resources)

Vendor/Type	**Full Text**	**Original Data**	**Directory & Network Data Which Identify Community Members to Each Other and/or the Public**
Enthusiast	0	4	0
Government & Education	0	18	3
Commercial Provider/Search Only	0	3	0
Commercial Provider/Portions of Data Restricted to Users Who Have Purchased Passwords.	0	10	0
Commercial "Public Service" Provider	2	3	12
Commercial Database Limited to Provider's Own Products	0	8	0
Commercial Database Offered At No Charge to Sell Enhanced or CD-ROM Versions of the Same Data, and/or Related Products.	0	4	0
Commercial Database Paid for by Advertising and/or Selling Right to Post Items on a Public Bulletin Board.	0	4	0

Table 5

Discussion. Tables 3 through 5 share striking similarities with the broader electronic databases discussed in Section A. In particular, they show:

C **Richness.** The sheer number and diversity of available databases is astonishing.

- **Diverse Suppliers.** The products listed in all three Tables are produced by government laboratories, private institutes, and commercial ventures. Researchers appear to use these sources interchangeably;

- **On-Line Versions of Print Media.** Academic journals and societies have rushed to make on-line versions of their journals available. This is exemplified by the fact that the American Physical Society, American Mathematical Society, American Chemical Society, and American Astronomical Society currently place *all* of their journals on-line. Most of the index and bibliographic products listed above are also extensions of preexisting print-based counterparts[*];

- **Self-Help.** Private publishers universally rely on passwords and/or contractual restrictions to limit access to, and republication of, their products[**];

- **Electronic Options.** Despite the greater technical difficulty of protecting CD-ROMs, they continue to be well-represented in the sample.

Example 5: A Large Nuclear Science Database[5] Since the late 1940s, the nuclear science community has struggled to reduce an exploding literature to a more manageable data set. Despite declining manpower and budgets, the Department of Energy ("DOE") continues to spend approximately $4 million per year to maintain, update, edit, and disseminate nuclear science databases.[***] Approximately $800,000 of this is spent to support a group at Lawrence Berkeley Laboratory ("LBL"), whose principal product is known as the *Isotope Table*. The *Isotope Table* is currently available as a book, on CD-ROM, and over the World Wide Web. The product includes over 160,000 published references and approximately 1.5 gigabytes of data.

Historically, nuclear database creators have never started "from scratch." For example, *Table of Isotopes* can trace its lineage to roughly half a dozen nuclear databases, many of which still exist. The LBL group has made extensive efforts to improve and extend these sources by adding new data, checking reported calculations, comparing different experiments to arrive at "best values," and deducing additional data not calculated by the original authors. *Table of Isotopes* is currently five years behind the literature on average.

Approximately one-half of the group's budget goes to improving their database so that it

---

[*] By way of example, the Berkeley Physics Department Web site reports that Inspec, MathSciNet, and Chemical Abstracts all existed on paper before their current, electronic incarnations. Inspec is over 100 years old.

[**] For example, the Engineering Library's Web site lists 47 of its 60 Web sites as "UC only" or "UCB only." Publisher Web sites were similarly restricted, although four offered their products on a "trial access basis."

[***] This figure does *not* include the actual work of reviewing articles, which is done on a volunteer basis throughout the world.

can support more advanced, relational searches; the balance is spent on disseminating their product over the Web and/or rearranging the data into new tables aimed at medicine and other non-traditional users.[*] DOE has not asserted any proprietary interest over the database. The LBL group is not worried about copying provided that proper attribution is given.

In addition to its public domain/Web-based version, *Table of Isotopes* is also available as a commercial book and CD-ROM. To protect against copying, the publisher has insisted on the following "self help" provisions:

- **Updates.** The group must supply new material annually, although the content of updates is discretionary. In practice, the group has concentrated on developing new tables aimed at non-traditional (and potentially lucrative) users such as medicine;

- **Additional Graphs.** The group must prepare copyrighted graphics that are (at least initially) superior to those found on DOE's Web site. This is an important selling point for commercial buyers who use the CD-ROM to prepare graphics for talks and presentations. The material also adds copyrighted content to an otherwise "public" product.

- **Additional Software.** The group must prepare additional software. This provides an additional selling point and copyrighted content not found at DOE's Web site.[**]

Although these enhancements are useful, the LBL group probably would have invested its resources differently if left to its own devices. In particular, they would have devoted more effort to updating and improving the underlying (but unprotected) database itself. This is a concrete example of how reliance on self-help solutions can distort investments compared to a hypothetical world in which all forms of intellectual property were identically protected by statute.

At the same time, the Berkeley group does not seem to view "self-help" as a significant bottleneck to new commercial projects.

---

[*] The Brookhaven National Laboratory has followed a similar path with respect to its related ENSDF database. Like Berkeley, Brookhaven has devoted extensive effort to (i) editing ENSDF's data, and (ii) improving ENSDF so that it can support advanced relational search engines. Brookhaven has also created a new version of ENSDF for use by medical workers. Finally, it is working to improve dissemination by upgrading its Web site and making the same data available on floppy disk and CD-ROM. *See*,National Research Council, *Bits of Power: Issues in Global Access to Scientific Data* (n.a.:1997) ("*Bits of Power"*) at pp. 205-6.

[**] Surprisingly, the private CD-ROM/book package competes successfully – and indeed seems to benefit from – its Web-based counterpart. In addition to the relatively minor enhancements required by the publisher, there seem to be intrinsic reasons for this. For example, books are often easier to use; searches conducted over the Web are not confidential; and CD-ROMs are "permanent" whereas data on the Web can potentially change or disappear without warning.

Example 6: Elsevier Science[6]

Elsevier Science ("ES") publishes (i) nearly 1,200 English-language scientific journals, (ii) a variety of highly specialized reference works, (iii) various bibliographies, abstracts, and reviews, and (iv) paper and electronic versions of the world's "most comprehensive interdisciplinary engineering database."* Virtually all of these materials are available both on-line and as CD-ROMs. ES's search software permits users to search multiple journals at once.

Although old print journals never had enough space to include full data sets, the advent of online journals has effectively removed this constraint. As a result, ES now requires authors to submit underlying data sets so that they can be "linked" to on-line journals.

ES routinely asks authors for the copyright to their work (including any underlying data), but will usually agree to accept a license instead. According to the company, there is currently no other way to manage reprint and reuse requests. The company does not ask for patent or database rights.**

Elsevier says that its non-scientific divisions have sometimes decided not to invest in new databases because of protection concerns. So far, however, this has not happened to any of ES's science projects. At most, database protection has been one issue among many.

To date, ES has collected only a "tiny" number of databases and has little experience with database issues. In analogy with its current reprint policy, the company (i) would probably not assert its copyright against authors who tried to make commercial products from their own previously-submitted databases, but (ii) probably would demand reasonable reprint fees from third parties who wanted to republish the data for commercial gain. The company has given little, if any, thought to compiling its own commercial products from authors' datasets. In *theory,* ES could assert its rights more aggressively in the future. Under this scenario, the company's large number of journals might then be leveraged into a corresponding dominance over databases.[7] So far, however, there is little indication that ES's disparate databases will ever be combined into a useful -- much less dominant -- commercial product.

Example 7: Biotechnology[8]

Bioinformatics. Finding commercially "interesting" genes is essentially a race to find subtle patterns in an enormous body of experimental data. (The task is often compared to that of prospectors looking for hints of gold in an otherwise featureless landscape.) The principal "raw data" needed to conduct academic and commercial biotech research is currently maintained in over

---

* Reed-Elsevier also publishes many non-scientific databases, including *"The Official Airline Guide."*

** ES points out that these policies are broadly similar to those of many other journal publishers, including Academic Press, the American Chemical Society, the American Institute of Physics, and the American Geophysical Union.

200 public sector databases scattered throughout the world. Virtually all of these Web sites are narrowly focused on the owner's research agenda. As a result, the system is often fragmented and redundant. From a computing perspective, many of the sites tend to be amateurish, underfunded and unstandardized.[9] This creates recurring difficulties for corporate users.*

The intersection between computer science and biology is known as "bioinformatics." Next generation bioinformatics systems will be designed to (i) convert diverse databases to a format that users can read, (ii) simultaneously search the Web's 200 + sites as if they were a single database, (iii) enhance existing text-based databases with "relational" links to make them more amenable to sophisticated searches, and (iv) create software search tools that are not only powerful but flexible enough to let researchers study the data in unanticipated ways.

GenBank. The best-known and most important public database is a National Institutes of Health ("NIH") Web site called GenBank. GenBank is one of three official locations in which researchers can "deposit" information about the precise order of "base-pairs" found in human DNA. The current Release 110.0 of GenBank contains over 3 million sequence records and includes more than 2 billion base pairs. More than 100,000 sequences from individual laboratories and high-throughput sequencing centers are added each month. Since it was founded in 1982, GenBank's size has doubled every 14 months.

Because of funding constraints, GenBank's capabilities are limited. For example:

- **Search tools** can only perform full text searches for written words. This is extremely unwieldy for most biology applications;

- **Editing and Comment** is limited to author annotations. No effort is made (i) to comment on related journal articles or (ii) to identify or resolve conflicts between data submitted by different researchers**; and

- **Updating** comments and sequences is virtually impossible.

These problems are not unique to GenBank. In recent years, several non-profit biotech databases have either closed or been threatened with closure. Commentators have complained that the community may have to "get by" with inadequate updating, editing, and annotations.[10]

---

* Because private biotech companies believe that submitting searches over the Web compromises security, each maintains internal copies of the 200+ public databases needed to conduct research and uses in-house software engineers to update them nightly. Since many on-line databases tend to change computing conventions abruptly, access often "crashes" without warning. The crashes cause recurring panics within corporate MIS departments.

** GenBank started out as a traditional database which tried to comment and add value to journal articles. GenBank converted to its current format because it could no longer keep up with the data.

Private Database Vendors. Beginning in the early 1990s, several firms began to offer private versions of a few databases to elite users willing to pay multimillion dollar license fees.* Initially, these databases were attractive because (i) they included large amounts of secret (*i.e.* proprietary) data, and (ii) they offered advanced bioinformatic search tools. Because public discovery was booming, the former advantage was short-lived. This has driven some firms to shift their emphasis to "the sale of new databases, software packages, and perhaps consulting."[11]

One early leader in the field (**Human Genome Sciences** or "HGS") started off by selling its proprietary database to a single research partner as part of a $125 million deal.[12] More recently, HGS has broadened its relationships and now plans to offer "market friendly" software packages ranging from "simple, low end packages for impoverished [academics to] tailor-made luxury items for drug companies." Raw data will be provided free of charge. [13]

HGS's principal rival, **Incyte,** originally charged licensees $15 - 20 million for access to its proprietary databases over a three year period.** Approximately 50 companies currently subscribe. Like HGS, Incyte focuses on software and database enhancements. According to Incyte's CFO,

> ". . . we think there's a huge information-based business growing from the pharmaceutical industry . . . This is not a small market segment that's going to be serviced by half a dozen companies. This is going to be a fairly large segment of service for a lot of companies, for everything from software and hardware companies to more biologically oriented companies and consulting firms that do systems integration or go in and design something specifically for a big drug firm."[14]

Incyte subscribers can currently buy the company's advanced "LifeSeq" relational databases with or without proprietary data. However, even non-proprietary databases have been cleaned and standardized to support Incyte's advanced search software.[15] Incyte also develops custom databases for individual clients; these are typically resold to other companies after an initial period of exclusivity.

HGS and Incyte have recently been joined by a third company called **Celera.** Celera's proposed human genome database will reportedly include (i) extensive proprietary human genome data, and (ii) a "value-added software and informatics system." Celera has not asked to share in the profits from any discoveries. Instead, it will offer its databases to users on a straight fee-for-service basis. Very large users will be able to purchase "dedicated systems."[16]

---

* These products were concentrated in particularly lucrative areas such as "expressed sequence tags" ("EST"s) or, more recently, gene sequences.

** The fee amounted to roughly 1% of a typical large pharmaceutical company's R&D costs.

HGS, Incyte, and Celera have many smaller rivals. These firms rarely sell proprietary information at all. Instead they concentrate on helping clients to manage their existing data in new and better ways.

Examples 8(a) - (h): Anecdotes and Profiles. The following examples are drawn from earlier descriptions of databases found in the literature[17]:

- **Example 8(a): Poisindex.** Poisindex is a CD-ROM product that links approximately 750,000 poisons to 775 management and treatment protocols. Approximately 200 clinicians from 20 countries participate in editing and selection. Poisindex also hires computer scientists to maintain its database and create search software. Poisindex is updated quarterly and sold by subscription.

- **Example 8 (b): MDL Drug Data Report.** This Reed-Elsevier CD-ROM database contains molecular structure and biology information for approximately 85,000 potential drug candidates. The data (i) is updated on a monthly basis from published reports, patent applications, and scientific papers, (ii) allows users to track clinical trials, (iii) comes with ISIS software that allows users to analyze the likely effects of modifying known drugs, and (iv) can be combined with the user's own data to create individualized research tools.

  MDL created the Drug Data Report (and seven other databases) because there was "an inadequate supply of scientific data for its ISIS software system." MDL offers a preferred fee for academic users.

- **Example 8(c): Visible Human.** This product consists of 10,000 "sliced" images of the human body and comes with software tools that include a navigator, bookmarks, and animation. Although originally developed from a database compiled by the US Government, Visible Human only became widely available after it had been commercialized. The government is still engaged in a major effort to update the underlying database.

- **Example 8(d): Derwent World Patents Index**. This database lists seven million inventions compiled from 13 million patent documents worldwide. Scientific journals and conference papers are also reviewed. The database is updated quarterly and is available on-line, as a CD-ROM, and in print.

- **Example 8(e): National Agricultural Database Library.** These animal husbandry databases are edited by a University of Wisconsin institute and are based on government publications submitted by Agricultural Extension Offices around the country. The database is chiefly used by educational institutions. It is available both on-line and as a CD-ROM.

C **Example 8(f): Materials Science.** A commercial gateway service known as the "Science and Technology Network" ("STN") provides access to "20 databases covering the physical and mechanical properties of thousands of materials as well as more than 100 factual and bibliographic databases." All of STN's databases can be searched simultaneously using a single set of sophisticated search tools.

STN's databases tend to be fairly permanent, but grow when new materials, conditions, and properties are measured. Users search STN's databases but are not allowed to download them. According to the NRC's *Bits of Power* study, many materials scientists believe that STN does not contain enough databases.

C **Example 8(g): Chemical Sciences.** Because of their long-standing ties to industry, chemists tend to provide a favorable environment for new commercial databases.[18] Private sector databases include The Registry of Toxic Effects (full text), Chemical Abstracts (bibliography), Derwent and STN International (patents), DETHERM (thermophysical properties), and SPECINFO (nuclear magnetic resonance and infrared spectra). Tabulations of evaluated data are also compiled by *The Journal of Physical and Chemical Reference Data.*

Publicly maintained databases include the Beilstein Institute (organic substances), Gmelin Institute (inorganic and organometallic substances), The National Institute of Standards and Technology (atomic species properties), and Cambridge University (structural data).

Chemists typically need to search multiple databases for any given task; sophisticated software search tools have been developed to do this. Most of the foregoing databases are available both on-line and as CD-ROMs.

• **Example 8(h): Geophysics and Meteorology.** Large databases in the earth sciences tend to be run almost exclusively by government agencies and clearinghouses. However, individual researchers frequently create smaller commercially valuable data sets in the course of writing papers. The American Geophysical Union may have further information on how such data is handled.[19]

C. Lessons Learned

The foregoing examples contain several recurring themes that need to be considered before attempting any reforms. These can be summarized as follows.

1. Protection Under Existing Law.

a. Compilations of Copyrighted Materials. Copyrighted information does not lose its protected status simply because it has been incorporated into a database. This fact is particularly important for so-called "full text" databases, which often consist *entirely* of copyrighted documents. Full-text databases have played a prominent role in several Examples.[*]

b. Copyrighted Enhancements. Databases are frequently sold together with advanced software as a single package. Since software is copyrightable (and often patentable), would-be copiers are faced with the choice of marketing a less capable product or else investing the resources needed to develop their own search tools. Copyrighted enhancements frequently appear in the Examples.[**]

2. "Self Help" Protection.

a. Bilateral Contract. The Examples include both custom databases prepared for a single customer and semicustom databases prepared for a relatively small community.[***] In the limit where only one customer wants to acquire a particular database, protection against third parties is virtually automatic.

More generally, existing law allows custom and semi-custom database owners to limit each customer's right to use and/or disclose the information to others. Such contracts are enforceable as trade secrets even where the underlying information does not qualify for statutory protection.[20] In practice, it is probably not feasible to negotiate and monitor more than a few dozen contracts at any one time. This limits dissemination to a comparatively small number of customers.

---

[*] *See* Examples 1 (commercial CD-ROMs), 2 (Web sampler), 4(a) (full text physics journals), 4(b) (full text engineering journals), 5 (copyrighted nuclear science graphs), and 6 (Elsevier Science full text journals).

[**] *See*, Examples 1 (commercial CD-ROMs), 2 (Web sampler), 5 (nuclear databases), 6 (Elsevier Science search engine), 7 (biotechnology databases), 8(a) (Poisindex software) and 8b (MDL Drug Database software).

[***] *See* Examples 3(a) (Dataquest semi-custom reports) and 7 (semi-custom databases in biotechnology).

b. Shrinkwrap Licenses. So-called "shrinkwrap" and "clickwrap" licenses can be used to bind an unlimited number of customers and are a ubiquitous feature of life on the Web.* Assuming that they could be enforced, most of these licenses would create protection comparable to, if not stronger, than that found in patents or copyright. The legal validity of such licenses is briefly discussed at Part II.B, below.

c. "Search-Only" and "Password Protected" Web Sites. The Examples include both "search-only"** and "password protected"*** Web Sites.[21] Such strategies are probably more viable in academia, where researchers tend to be less concerned about security. Private sector corporations, on the other hand, tend to avoid using the Web because (i) communications (including search requests) are not secure, and (ii) Web Sites can change (or disappear) overnight. For this reason, corporations insist on using CD-ROMs wherever possible.

d. Updating. Seventy-three percent of the CD-ROM products listed in Table 1 were regularly updated on an annual, quarterly, or monthly basis.[22] Regular updating was also a recurring feature of various other Examples described above.****

The consumer preference for updating is understandable where (i) data change quickly, or (ii) even a small chance of error could compromise large investments of time, labor, and capital. A good example of an industry where both factors apply is biotechnology.

e. Editing and Enhancements. The central importance of editing and enhancements in the sciences recurs throughout the Examples.***** Given the premium which science and engineering place on editing and enhanced databases, similar features would probably exist even without the threat of copying. This does not change the fact that editing and enhancements promote self-help protection by providing (i) added copyright content, and (ii) occasions for frequent updating.******

---

* *See*, Examples 1 (Internet sampler), 2 (CD-ROM sampler), 4(a) (physics CD-ROMs), 4(b) (engineering CD-ROMs), and 4 (c) (engineering Web sites).

** *See* Examples 1 (Internet sampler) and 8(f) (materials science database).

*** *See* Examples 1 (Internet sampler), 4(a) (physics journals), and 4(b) (engineering resources).

**** *See*, Examples 3(b) (Info-Trac), 4(a) - (c) (journals, indexes, and bibliographies), 5 (nuclear science), 6 (Elsevier Science), and 7 (biotechnology).

***** *See* Examples 5 (nuclear physics), 7 (biotechnology), 8(a) (Poisindex), and 8(b) (animal husbandry).

****** *See*, Example 5 (updating of nuclear physics databases to reflect improved data), 7 (biotechnology) and 8(f) (updating of materials science databases to reflect improved data).

3. Unprotected Products.

a. Public Domain. A significant number of products are sold without any protection at all, sometimes for comparatively high prices. One suggestion invokes market imperfections as an explanation. Under this scenario, large vendors who can afford to circulate catalogs are able to make a profit on relatively obscure titles even if only a few customers purchase them. Would-be copiers are too small to reach these same customers and therefore do not compete.

b. Spinoffs. In the paradigmatic *Feist* case, plaintiff created a telephone directory as an essentially cost-free spinoff of providing telephone service. In Example 8(b), a provider created eight new databases in order to promote sales of its core product (software).[*] In both cases, the provider could reasonably expect to recover its investment whether or not its databases were later copied.

These examples show that providers will continue to produce some databases regardless of whether they are protected.[**]

---

[*] This is also a popular strategy for Web-based businesses. *See* Example 1.

[**] It might be argued that there is no reason to enact legislation that "encourages" cost-free databases because such "spinoffs" will exist whether or not they are protected. This argument ignores the role of price signals in achieving economic efficiency. If industry members are not allowed to recapture the value of spinoffs, the underlying product will be more expensive (and less used) than it should be. Suzanne Scotchmer, "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 29-41 (Winter 1991).

Part II: Existing and Proposed Law

A. The Limits of Federal Copyright Protection.

For most of the Twentieth Century, the extent to which databases were or were not protected was uncertain. The federal copyright and patent statutes seemed to be exclusive. Congress had not afforded database protection, the argument went (and still goes): Therefore, there shouldn't be any.

The *INS* Case. The legal history of database protection in the US begins with the Supreme Court's 1918 decision in *International News Service v. Associated Press* ("*INS*").[23] The case involved a wire service whose employees rewrote its rival's published dispatches and then sold them as its own. On appeal, the Supreme Court was asked to decide whether the copyright act's policy in favor of putting facts into the public domain created an absolute right to engage in such practices.

The Supreme Court decided that it did not. Instead, it drew a distinction between the *public's* right to information (which was protected by copyright) and a business competitor's right (which was not).[24] The Court's policy analysis was starkly modern in its use of economic reasoning:

> "Indeed, it is one of the most obvious results of defendant's theory is that, by permitting indiscriminate publication by anybody and everybody for purposes of profit in competition with the news-gatherer, it would render publication profitless, or so little profitable as in effect to cut off the service by rendering the cost prohibitive in comparison with the return."[25]

Although succeeding courts wrote a handful of opinions applying *INS*'s reasoning to new facts, they did little to clarify the Supreme Court's attempted distinction between "fair" and "unfair" uses of public knowledge. Instead, the case remained in a kind of legal limbo. In the words of one commentator:

> "Having been born over the objection of powerful dissents authored by Justices Holmes and Brandeis and thereafter subjected to the disapproval of Judge Learned Hand, it is not surprising that the *INS* case was 'often confined strictly to its facts.' The upshot has been recognition of a claim where the defendant's commercial exploitation was likely to destroy its value but otherwise to allow the defendant to compete notwithstanding the advantage gained by use of the plaintiff's work."[26]

The *Feist*[*] Case. The modern era of database law opened with the Supreme Court's 1991 decision in *Feist Publications, Inc. v. Rural Telephone Service Co.*[27] The case stemmed from a publisher's attempt to copy a local telephone company's printed directories.[28] Starting from the twin propositions that "facts are not copyrightable" but that "compilations generally are,"[29] the Court explained that creating a telephone book in the usual format lacked the "minimal degree of creativity" required for copyright protection under the US Constitution.[30] The Court then effectively decided the case a second time by concluding that earlier cases which had extended the copyright statute to works created by "sweat of the brow" rather than creativity had been wrongly decided.[31] However, the Court stopped short of overruling *INS*. Instead, it only distinguished the case by saying that it had been decided "on noncopyright grounds that are not relevant here."[32]

Copyright Protection After *Feist*. Since 1991, approximately a dozen courts have analyzed and elaborated on the principles announced in *Feist*. In doing so, they have frequently found that the compiler's choice and arrangement of data can be sufficiently creative to trigger copyright protection. *Warren Publishing, Inc. v. Microdos Data Corp.*, 52 F.3d 950 (11th Cir. 1995) (taking facts from an "external universe of existing material" and arranging them according to an idiosyncratic list of "principal communities" was sufficiently creative to qualify as a copyrighted compilation)[33] ; *Key Publications, Inc. v. Chinatown Today Publishing Enterprises, Inc*., (selection of businesses to be included in directory "was in no sense mechanical, but involved creativity . . . in deciding which categories to include and under what name")[34]; *but see, Bellsouth Advertising & Publishing Corp. v. Donnelly Information Publishing, Inc.*, (fact that company's telephone directory limited entries to subscribers living within a certain region on or before a particular closing date did not satisfy *Feist*)[35]. The only significant qualification seems to be that such creativity must involve the *arrangement* of data and not the discovery of information itself. *Bellsouth, supra*, (copyright act "affords no shelter to the resourceful, efficient, or creative collector").[36]

From the standpoint of science, the most important *post-Feist* development involves case law suggesting that compilers who apply judgment to their data also qualify for copyright protection. *Mason v. Montgomery Data, Inc.*, (plaintiff applied "discretion" to task of selecting, interpreting, and reconciling inconsistencies among sources)[37]; *Nester's Map & Guide Corp. v. Hagstrom Map Co.*, (author recommended best ways to find particular buildings and approximated street addresses so that they would be easier to remember)[38]; *CCC Information Services, Inc. v. MacLean Hunter Market Reports, Inc.* (price estimates based on "professional judgment and expertise" rather than "reports of historical prices" or "mechanical derivations of historical prices or other data" were copyrightable).[39]

---

[*] One confusing aspect of *Feist* is that many commentators who disagree with the Court's reasoning nevertheless support its final ruling. For example, Tyson & Sherry argue that telephone book data should not be protected because it is generated "with no additional effort" in the course of operating a publicly-sanctioned monopoly. Laura D'Andrea Tyson & Edward Sherry, *Statutory Protection for Databases: Economic and Public Policy Issues* (1997) (report commissioned by the Information Industry Association). In narrowly legal terms, the same result could also be reached by arguing that firms which exercise "monopoly power" in one market should not use it to obtain an "unfair" cost advantage elsewhere.

The fact that many courts have been willing to find "creativity" in the way that databases are arranged does *not* mean that the data itself is protected. If "free-riders" are willing to take the time and trouble to select from and rearrange copyrighted databases they remain free to do so. *See, e.g., Warren Publishing, supra* ("content of datafields" was "merely fact[]" and not copyrightable)[40]; *Skinder-Strauss Associates v. Massachusetts Continuing Legal Ed., Inc.,* (bare fact that defendant copied information from plaintiff's directory did not establish copyright violation)[41]; *Cable News Network, Inc. v. Video Monitoring Services of America, Inc.,* (copyright in news broadcast only extended to compilation as a whole; individual news segments remained "factual in nature" and unprotected)[42].

Unfair Competition After *Feist*. Given *Feist's* extensive criticism of the "sweat of the brow" doctrine, it would have been reasonable to think that *INS*-type unfair competition claims had no further validity. However, this turns out not to be true. Instead, the prestigious Second Circuit Court of Appeals declared in *National Basketball Assn. v. Motorola, Inc.*[43] ("*NBA*") that the core situation addressed by *INS* -- the so-called "hot news" cases -- remained good law. Because the Second Circuit carefully restricted its discussion to "time sensitive" information, the decision does not currently include databases.[*] Nevertheless, the fact that the Second Circuit continues to take unfair competition seriously suggests that the doctrine may have a future. A detailed discussion of the *NBA* case and its potential extension to databases can be found at Part III, below.

Implications for Science. To date, no court has applied *Feist* to scientific databases. Given the relatively low standards required to find "creativity" under the case law, it seems clear that (i) extensive editing judgments, (ii) attempts to choose "best values," or (iii) enhanced "relational" search capabilities would all qualify for "compilation" protection under the copyright laws. The smaller, less elaborate data sets submitted in connection with individual papers would probably also qualify, although this presents a closer question.[**] At least one publisher (Elsevier Science) appears to be operating under the assumption that they do.[44]

If copyright applies, it still would not prevent would-be copiers from extracting and rearranging data from a scientific database. The "minimum" amount of creativity that such copiers would have to show is an open question, but might be quite minimal.[***] In any case, it seems safe to say that the traditional scientific practice of compiling new generation databases

---

[*] In theory, database owners could argue that database *updates* are "time-sensitive" in an economic sense and should therefore be protected. This would require a semantic stretch beyond anything in *NBA* itself.

[**] The obvious counter argument is that many scientific databases follow conventions that leave little room for creativity. For example, spectra almost always show frequency on one axis and amplitude on the other. A better argument might be that the experimenter's choice of which data to present still reflects creative choices. Even this argument might not be enough for human genome sequencing or other areas of routinized inquiry.

[***] The EU's Directive of Databases has suggested that existing databases could even be "rearranged electronically . . . to produce a database of identical content which, however, does not infringe any copyright in the arrangement of [the] database." *Directive* at ¶ 38.

from earlier ones remains viable. *Cf., Sinai v. California Bureau of Automotive Repair* (defendant was free to compile its own manual from data listed in copyrighted chart).[45]

B. State Contract Law.

Despite their ubiquity, the effectiveness of so-called "shrinkwrap" or "clickwrap" licenses remains unclear. Traditionally, courts have often been willing to look past the fiction that purchasing an article constitutes "agreement" to a license, particularly where the license is one-sided. Recent case law has held that shrinkwrap licenses can indeed be used to obtain greater rights than those obtainable through copyright. *Pro-CD, Inc. v. Zeidenberg*. (enforcing shrinkwrap license restrictions protecting telephone listings database against copying)[46]. Nevertheless, the doctrine's outer bounds remain unclear. *Vault Corp. v. Quaid Software Ltd.*, (federal law invalidated contractual restriction against decompiling computer program)[47]

Implications for Science. Existing case law suggests that "shrinkwrap" licenses are a viable strategy for extending database rights beyond copyright and patent law. The outer limits of such protection will ultimately be set by the courts' willingness to find overreaching provisions "unconscionable" or otherwise "contrary to public policy."

C. The European Union Directive.

Databases have traditionally received substantial protection in the U.K., Ireland, the Netherlands, and the Nordic countries. Until recently, however, they received much less protection elsewhere in Europe. Beginning in the late 1980s, the European Union (EU) began studying database protection as part of a larger project to harmonize member states' copyright laws. Although initial proposals were relatively moderate, calls for protection grew steadily stronger over time. In March 1996, the EU Council issued a "Directive on the Legal Protection of Databases."[48]

By its terms, the EU Directive applies to any "collection of independent works, data, or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means."[49] The Directive protects such works against "temporary or permanent reproduction,"[50] "adaption" or "alteration,"[51] or "distribution to the public."[52] However, these protections do not apply unless a "substantial part" of the database, "evaluated qualitatively and/or quantitatively," has been copied.[53]

The EU Directive provides protection for a period of 15 years.[54] However, this period can be indefinitely extended if the "accumulation of successive additions, deletions, or alterations" amount to "substantial new investments." This extension would extend to the database as a whole and not just to "new" components.[55]

To American eyes, the most striking aspect of the Directive is its refusal to extend protection to citizens of countries that do not adopt the EU's standards. Formally, this is implemented by Art. 11, ¶ 3, which gives the Council discretion to withhold database protection from "databases made in third countries . . ." The Directive's preamble makes the underlying threat:

> "[T]he right to prevent unauthorized extraction and/or re-utilization in respect of a database should apply to databases whose makers are nationals or habitual residents of third countries or to those produced by legal persons not established in a Member State, within the meaning of the Treaty, only if such third countries offer comparable protection to databases produced by nationals of a Member State or persons who have their habitual residence in the territory of the Community."[56]

Implications for Science. From the standpoint of the scientific community, one of the most important aspects of the EU Directive is found in Article 6, which gives member states the option to exempt copying

> ". . . for the sole purpose of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved."[57]

A related provision preserves member state exceptions "traditionally authorized under national law."[58] This provision presumably includes European equivalents of the American "fair use" doctrine, which permits limited copying for scholarship and research.[59]

D. Proposed Legislation and Suggested Reforms.

1. The WIPO Treaty.

The August 30, 1997 Draft. Shortly after issuing its Directive, the EU asked the World International Property Organization to consider a worldwide database treaty based on the European model. After preliminary discussions involving the US, WIPO published a draft version on August 10, 1996.[60]

By its terms, the WIPO draft would have defined "databases" to include

> " . . . a collection of independent works, data or other materials arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means."[61]

The heart of the treaty would have required member countries to adopt legislation granting

database owners an "exclusive right" to prevent "the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium" without their permission.[62]

Unlike the EU Directive, the WIPO treaty would have restricted the right of member states to enact exemptions for scientific research, either directly or through "fair use" provisions:

> "Contracting Parties may, in their national legislation, provide exceptions to or limitations of the rights provided in this Treaty in certain cases that do not conflict with the normal exploitation of the database and do not unreasonably prejudice the legitimate interests of the rightholder."[63]

Implications for Science. Far from resisting European efforts to implement database protection, the initial US reaction was to *extend* the EU Directive's coverage by requesting a 25 year period. The WIPO draft left the question open.[64] As previously noted, the WIPO draft also would have sharply reduced member nations' right to exempt scientific research.

The WIPO Treaty Derailed. Although the Clinton Administration originally backed the proposed treaty, support was split after US scientists and developing nations protested in late 1996. Talks in Geneva were finally derailed roughly one week before the treaty was to have been completed.[65] WIPO is still examining database protection.

2. Proposed Congressional Legislation.*

WIPO's database protection provisions could not have been implemented without domestic legislation. Such legislation was first introduced in May 1996. Later, a similar bill (HR 2652/S2291)** was offered as an amendment to an intellectual property reform package that ultimately became known as the "Digital Millennium Copyright Act of 1998."[66] HR 2652 was dropped in Conference Committee (largely because of protests from the science and engineering community) but will be reintroduced this year. The bill can be broadly summarized as an "implementation statute" designed to meet the EU's demands. HR 2652 frequently paraphrases or incorporates the Directive verbatim.

As written, HR 2652 would have protected "information that has been collected and has

---

* The question of whether the Constitution allows Congress to pass European-style database legislation is outside the scope of this report. For a list of possible problems, *see* U.S. Copyright Office, *Report on Legal Protection of Databases* (Aug. 1997).

** A copy of HR 2652 is attached as Exhibit 9 hereto.

been organized for the purpose of bringing discrete items of information together in one place or through one source so that users may access them."[67] The bill would have imposed liability on

> "Any person who extracts, or uses in commerce, all or a substantial part, measured either quantitatively or qualitatively, 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 . . . for a product or service that incorporates that collection of information and is offered or intended to be offered for sale or otherwise in commerce by that person . . ."[68]

Despite this provision, users who extracted individual facts or "insubstantial parts" of databases would not be liable.[69]

HR 2652 also would have provided additional exemptions for the following conduct:

- Users could gather the same information independently and even use protected databases to verify the accuracy of their research[70];
- Users could copy material for "nonprofit educational, scientific, or research purposes in a manner that does not harm the actual or potential market for the product or service . . ."[71] ; and
- Users could use information "for the sole purpose of news reporting" so long as they did not copy time-sensitive "hot news."[72]

The government and its employees would not have been allowed to claim protection under the statute. However, educational institutions would have been.[73]

HR 2652's principal enforcement mechanism would have depended on civil suits for damages, lost profits, injunctions, and treble damages in appropriate circumstances.[74] Criminal penalties would also have been available.[75]

Implications for Science. Like the WIPO proposal, HR 2652 would have extended protection for databases *beyond* the EU's demands by (i) extending the term of protection from 15 to 25 years, and (ii) failing to take full advantage of the EU Directive's exemption for non-profit copying and/or fair use for teaching and scientific research.

3. State Legislation.

Because shrinkwrap licenses are creatures of contract law, state legislatures have considerable power to limit or extend them by statute. In the interests of commerce, however, they have generally worked together to enact nationally uniform laws. One such draft currently

under discussion would amend the Uniform Commercial Code ("UCC") to codify and change the law of licenses. Although the proposed statute is complex, the most relevant section for present purposes involves enforceability:

> "If a court as a matter of law finds the contract or any term of the contract to have been unconscionable or contrary to public policies relating to innovation, competition, and free expression at the time it was made, the court may refuse to enforce the contract . . ."[76]

"Reporter's Notes" to the proposed statute explain that it is designed to acknowledge the fact that (i) "State laws, including the UCC, cannot alter or create federal law,"[77] (ii) that public policy with respect to "innovation, competition, and free expression policy" may render contracts invalid even when federal law is silent,[78] and (iii) that the UCC "take[s] no position" on "a general federal policy question."[79] On the other hand, the provision reaffirms the traditional rule that "private parties may have sound commercial reasons for contracting for limitations on use . . ."[80] even where those restrictions exceed the bounds of copyright.[81] Such vague and conflicting comments will provide little guidance to courts trying to determine whether a particular shrinkwrap contract satisfies federal law. The expectation seems to be that the courts will eventually reach some sort of consensus.

More concretely, the Reporter suggests that license terms which prevent customers from providing access to multiple users, using data for commercial purposes, or modifying a database's content "would *in most cases* be enforceable."[82] Assuming that courts heeded this advice, the last two restrictions could potentially eliminate the now-common practice of creating new databases from old ones.

4. Academic Proposals.

One of the most thorough reviews of the case law and recent history of database protection is found in a recent law review article by Reichman & Samuelson (hereinafter "Reichman & Samuelson").[83] The article concludes with two proposals: (i) an enhanced version of existing unfair competition law similar to, but more refined than that announced in the Second Circuit's *NBA* decision, or alternatively (ii) a "preferred solution" in which an industry-based "collection agency" would set "baseline" license fees. A detailed discussion and evaluation of these proposals is presented at Part III, below.

Part III: Policy Implications

A. Is There Room for Improvement?

Statutory protection should not be extended to databases lightly or for no reason at all. Potentially, the best reason to enact protection would be to encourage investors to create new databases that do not currently exist.* The chances that such an incentive would actually work is discussed as **Issue 1**. A second, weaker argument for statutory protection – that the existing system of self-help tends to create excessive secrecy and/or diverts investment away from databases into ancillary features – is discussed at **Issue 2**.

Issue 1: *If statutory protection were enacted, what new products would be created?*

Part I demonstrates that many types of databases can and do flourish in today's world. Traditionally, it has been argued that expanding these "islands" of legal protection would lead to more and better products. Human ingenuity being what it is, the Committee should not dismiss this possibility lightly. Nevertheless, the interviews and research conducted for this report did not find a single instance in which a commercial publisher decided not to start a project because it lacked statutory protection. At most, protection was one concern among many – and cost concerns unrelated to potential copying usually ended up "driving" the decision.**

The fact that databases have managed to prosper without statutory protection is sufficiently counterintuitive to ask whether economic theory can account for it. One possible clue is that most of the databases described in this report have existed for many years, so that the providers have long since recouped their initial "startup" investment. A second clue is that virtually *all* databases began as compilations of still earlier products. The assumption that database suppliers invariably face large "startup" costs needs careful examination.

But if database providers do not have a large up-front investment to protect, what *are* they investing in? The answer is updates, improvements, and extensions of their existing product. After all, database providers *already* "compete against their own product" because consumers can almost always continue to use last year's CD-ROM rather than buy a new one. A third party's

---

* This policy-oriented approach to the problem necessarily ignores justice-based appeals that creators "should" be compensated. Suffice to say here that strong normative arguments exist *against* rewarding inventors who knew in advance that certain types of activity would not be compensated.

** Karen Hunter of Elsevier Science did report that her company had turned down non-scientific database products because it was afraid of copying. Furthermore, it is possible and even likely that counterexamples in science and engineering could be found if a more systematic survey were conducted. The apparent rarity of such counterexamples is nevertheless striking.

decision to make *additional* copies of last year's CD-ROM is unwelcome, but conceptually similar.

If the foregoing observations are correct, the only new products that protection is likely to elicit would be large databases that cannot be assembled from precursors. Whether such projects actually exist is an open question.

Issue 2: *Has the absence of statutory protection resulted in economic distortions and, if so, how serious are they?*

a. Pathologies of Self – Help: Limited Access & Secrecy. One of the most effective forms of self-help relies on bilateral licenses in which users promise to preserve the database's secrecy. This method of protecting intellectual property has a long history and is not just confined to databases. The fact that some of today's firms still choose secrecy over patents shows that bilateral contracts can be a very powerful form of protection.

Secrecy is most effective when the number of authorized users is small because large numbers of customers increase the risk of leaks. For this reason, owners who rely on secrecy tend to concentrate on their most lucrative customers while foregoing other, less profitable transactions. Since the lost transactions would have made both seller and buyer better off, such conduct is socially inefficient.

Ideally, database owners could use other forms of self-help instead of secrecy. However, such strategies are usually not as reliable as secrecy. Particularly in the case of very valuable products, the owner may be afraid to use them.

Statutory protection offers a way out if (and only if) it is a reliable substitute for secrecy. In such instances, owners will typically try to charge their "core" customers a high price while offering lower, "preferred" rates to others (e.g., academic scientists). This so-called "discriminatory pricing" model[84] unambiguously improves efficiency by increasing the number of transactions between willing buyers and sellers in the economy.

Example 7 (biotechnology) shows how secrecy can end up restricting databases to a handful of firms willing to pay multimillion dollar license fees. But would statutory protection have led to a different result? In the case of biotechnology, the existence of customers willing to pay millions of dollars almost certainly would have led to cheating if the same database had simultaneously been offered to academic researchers at "affordable" prices. Nevertheless, statutory protection could still be viable under other, less extreme situations. The Committee should ask witnesses whether such situations actually exist anywhere in the sciences.

b. Benign Self-Help: Editing and Enhancements. This report has shown that updating, editing, bundling with advanced software, and other enhancements are all popular forms of self-help. The danger is that the need for self-help will encourage database owners to over-

invest in such features. If so, this distortion will be paid for by a corresponding under-investment in the database itself. A relatively mild instance of this appears to have occurred in connection with the privately published nuclear database discussed in Example 5.

In economic terms, the most effective self-help perspective requires providers to supply the enhancements that consumers want most. Since this is the same solution as that found in a free market, distortions should be minimal in most cases. This conclusion might *not* hold, however, if the total enhancements needed to effectuate self-help exceeded the amount that consumers would demand in a free market. In that case, providers would begin to over-invest in enhancements at the expense of underlying databases.

In the end, the issue of whether self-help introduces distortions depends on the empirical question of how many enhancements consumers would demand in a free market. However, this report has shown that free market demand for updating, editing, and other enhancements in the sciences is likely to be very strong. For this reason, broadly comparable activities are likely to go on *regardless* of whether statutory protections are enacted. The fact that self-help might produce unwanted distortions in some other, theoretical world hardly matters.

c. Summing Up. Most forms of self-help do not require secrecy and cause few distortions. A worrisome exception – bilateral contracts and secrecy – occurs in biotechnology but does not seem to be widespread elsewhere in the sciences. Even in biotechnology, there is little indication that database owners would actually give up secrecy if statutory protection became available.

B. Pitfalls and Drawbacks

Issue No. 3: *Will protecting first-generation databases discourage the creation of subsequent products?*

Scholars have pointed out that, under certain circumstances, protecting the rights of earlier innovators can discourage subsequent innovation. This has been called the "tragedy of the anticommons."[85] In the database context, the theory argues that giving "first generation" database owners the right to demand compensation (i) encourages "first generation" products, but (ii) raises the cost of producing later generation products ever afterward. If effect (ii) is larger than effect (i), statutory protection could actually end up *reducing* the total number of databases produced over time.

The tragedy of the anticommons will not occur if the creators of second-generation databases are allowed to negotiate licenses with existing database owners in advance.[86] The reason is that the existing owners can only earn licensing revenues if later-generation databases are actually created and sold. For this reason, the owners will always set their fees so that new projects to remain profitable.

The tragedy of the anticommons might still occur under two circumstances. First, many databases are created on a not-for-profit basis, in which case there will be no revenue stream to share with the first generation owner. In principle, this is not a problem because the first generation owner can still make a profit by hiring someone else to do the work commercially. In the narrowly specialized world of the sciences, however, such people may not exist.

The second circumstance in which the tragedy of the anticommons might still occur involves databases that are "public goods" – *i.e.*, cannot be sold at a profit even though they benefit society as a whole. In theory, the government should be prepared to pay whatever license fees are needed to produce these products just like any other cost of production. In practice, however, governments have often found it politically difficult to purchase intellectual property from the private sector. There is therefore no guarantee that such expenditures would be made in the future.

Issue No. 4: *Should statutory protection include exemptions for socially useful copying?*

a. Exemptions for Non-Commercial Use. Within the scientific community, much of the debate over database protection has centered on whether there should be exemptions for research.[*] The standard economist's objection to such exemptions is that they penalize the database owner. If research truly benefits society, the argument runs**,** it should be paid for out of general tax revenues. Conversely, forcing the database owner to give up part of his rights unfairly puts society's burden on a single individual.

On the other hand, the fact that society *should* use tax money to purchase databases for its researchers is no guarantee that it will. While this is essentially a political judgment, mainstream economics' so-called "theory of the second best" teaches that enacting *part* of a socially optimal plan is sometimes worse than doing nothing at all.[**]

The idea that businesses should make in-kind subsidies to "worthwhile" activities continues to exert a powerful hold over the American imagination. For example, many doctors donate their labor to charity. The appeal is particularly strong in the present situation. Having asked for unprecedented protection, the argument runs, database providers should not complain if they end up receiving slightly less than the entire pie.

Finally, society may decide that non-commercial databases are valuable in their own right and need to be protected. This question is addressed at **Issue 7**, below.

---

[*] *See, e.g.*, Andrew Lawler, "Database Access Fight Heats Up," *Science* (November 15 1996); *see also, Bits of Power, supra*, p. 171 (recommending that "fair use"-type provisions be included in any future database legislation).

[**] *See, e.g.*, Walter Nicholson, *Microeconomic Theory: Basic Principles and Extensions* (6th ed. 1995) at pp. 568 - 69.

b. A "Fair Use" Exemption for "Honest Copying" Most commentators assume that new databases are created (i) by gathering information *de novo*, or else (ii) by paying for the right to use someone else's database as a starting point. However, the Examples found in this report suggest that neither model has been particularly important in the past. Instead, scientists have normally used earlier databases without payment to create fundamentally different products. Since such behavior does not fit the normal "free rider" stereotype very well, it is worth asking whether it provides benefits and whether future reforms should try to retain it.[*]

From the economic perspective, the answer probably depends on how much additional protection databases need. If high levels of protection are needed to protect investors from free riders, the law should recognize few if any defenses to copying. But if the existing world of "self-help" is only slightly inadequate, statutory reform will have to include broad "fair use"-type defenses in order to avoid over-protection. Creating a "safe harbor" exemption for traditional scientific practices is one way to do this.[**]

In addition to such economic considerations, there are also sound legal arguments why some sort of fair use exemption should protect users who genuinely try to improve what they copy. The reason is that copyright *already* protects authors who take facts from existing works: Refusing to implement a similar fair use exemption for database copying would lead to an anomalous situation in which database protection actually exceeded that of copyright. The importance of maintaining the traditional distinctions between copyright and lesser forms of intellectual property are further discussed at **Issue 7**, below.

Issue 5: *Would increased rights allow database providers to charge higher prices within individual "niche" markets?*

Many observers argue that the sciences contain large numbers of "niche" markets, each of which is served by only one or two providers. The result, they claim, is a tendency towards price gouging that will only get worse if additional database protection is enacted. The usual economic counterargument is that existing providers cannot set prices too high without attracting competitors. For this reason, logic suggests that prices should remain at or near competitive levels even in single provider markets.

A more sophisticated argument suggests that the investment needed to enter a particular "niche" may be nearly as large as the market itself. In such cases, the entire market will never generate enough revenues for more than one or two firms to recover their investment. Since this

---

[*] An additional difficulty would be encountered during the transition period which followed any reform. This is because owners of pre-reform databases would receive full protection even though they had not paid for their own "headstarts" under the old system.

[**] Some proposed legislation suggests that copying should be permitted where the "new" database serves a different market than the first one. This is another kind of "honest copying" exemption.

fact deters would-be competitors from entering the market, existing providers can raise prices without fear of entry. Statutory protection would make this problem worse by adding to would-be competitors' initial startup costs.

The question of whether niche markets actually exist will ultimately have to be settled empirically. So far, however, most studies have confined themselves to counting the number of existing competitors in each market – thereby ignoring the crucial role of potential entrants. More definitive studies will have to look at entry costs and/or evidence of abnormal returns to capital. In the meantime, suggestive evidence that niche markets may exist comes from the fact that a closely related industry – scientific journals – has recently been accused of price gouging.[*] Inverting the normal economic argument, one could say that the existence of high prices in this area proves that entry is difficult. Statutory reform would make the resemblance between journals and scientific databases even closer than it is today.

Given the present state of the evidence, the Committee should be careful to ask witnesses for concrete examples of "price gouging." If such practices turned out to be widespread, they would constitute a strong argument against extending protection still further.

Issue 6: *Could statutory protection damage science by inadvertently "privatizing" its databases?*

Part I of this report shows that different branches of science have created database communities that range from "all public" to "all private" with every conceivable mixture in between. On the other hand, **Issue 3** has pointed out that protecting existing products can be a disincentive to further database creation (the anticommons problem) unless firms are able to buy licenses from one another. Since public databases do not generate the revenues needed to pay license fees, statutory protection imposes substantial (albeit inadvertent) pressures to privatize.

One problem with privatization is that most scientific databases are "public goods" that require government support. In theory, "privatized databases" could continue to receive such support in the form of government subsidies or grants. In practice, government may lack the political will to do this.[**]

Harder to quantify, but no less important, are the likely effects of privatization on information exchanges between scientists. As one witness said, "This"ll make it even harder for

---

[*] Journal prices rose 115% between 1986 and 1994. A leading study commissioned by the Association of Research Libraries blamed the increases on an "imperfect, monopoly-like marketplace" controlled by a small group of publishers. See, e.g., Gary Taubes, "Electronic Preprints Point the Way to 'Author Empowerment,'" *Science* (Feb. 9, 1996).

[**] *Cf., Bits of Power, supra,* at 114 (criticizing economic argument in favor of having researchers pay for databases from their individual research budgets as politically unsustainable).

me to give stuff away for free."[*] Similarly, there are already complaints that the EU Directive has made European scientists more reluctant to share data with their US collaborators.[**] Enacting a database statute could cause similar problems within the US itself.

Issue 7: *How would database protection interact with other forms of intellectual property protection?*

The definition of databases as "compilations of information" is troubling because practically anything – including *Gone With The Wind* – can be described as "a compilation of information." In the words of one commentator, "no abstract definition of a database will give us a bright line border between databases and non-database works."[87] This has prompted some scholars to supplement the definition by listing products that are *not* "databases."[88] Given the explosion of new forms of intellectual property, such negative definitions seem doomed to failure.

The Committee should recognize that phrases like "compilations of information" continue to be used because they summarize basic attributes that people want to protect. The fact that these attributes can be found in almost *all* intellectual property only shows that most of the arguments for database protection are very general. If the Committee ultimately agrees with these arguments, it should logically be prepared to advocate the same (or greater) protection for all other forms of intellectual property. Otherwise, the most general type of intellectual property (database protection) could end up becoming more desirable than the narrowly-defined categories traditionally thought to merit heightened protection under the copyright and patent statutes.

Some scholars have tried to limit databases to a special category of products for which "*sui generis*" laws can be written. This approach is unnecessary if database protection is simply thought of as the default choice for products that do not meet the relatively high standards of copyright or patent protection. Conversely, the Committee should be deeply suspicious of any proposal that would afford database owners *any* right that is not simultaneously available to copyright or patent owners.

---

[*] Interview with Thomas Slezak (bioinformatics expert), Exhibit 6; *see also,* Interview with Karen Hunter, Exhibit 5.

[**] Interview with Karen Hunter, Exhibit 6; *see also* Eliot Marshall, "Please Pass the Data," *Science*, 276:1961 (June 27, 1997) (reporting "recent pressure from [the EU] to give industry first crack at any genome data").

C. Threats From The European Union

Issue 8: *Does the EU's position on databases change the foregoing analysis?*

The principal reason for the EU Directive is that stronger incentives would encourage European companies to create more databases.[89] The EU's threat to leave American databases unprotected in Europe if the US does not pass reciprocal legislation also appears to have been motivated by the fact that databases are a worldwide market which require a consistent set of rules.

The fact that the threats contained in the EU Directive did not take effect immediately suggests that they may have been intended, at least in part, as a bargaining position. While it is true that some observers have called the EU Directive a hunting license to copy unprotected US databases, others have remained skeptical. Last year, *Science* reported that "observers on both sides of the Atlantic doubt that Europeans will do so because of fears that such a move could spark a trade war."[*]

If the US decides that EU-style database protection is not in its best interests, it should ask the Europeans to negotiate. If this fails, the US will have to decide whether its larger interests require it to enact EU-style legislation anyway. But there is no reason not to try.

---

[*] Andrew Lawler (ed.), "EU Database Directive Raises Hackles," *Science*, 279:165 (Jan. 9, 1998).

Part IV: Available Policy Tools

If the Committee finds that existing law needs to be reformed, it must next consider what tools are available. This section summarizes the various strategies found in existing or proposed database protection laws (*see*, Part II above) and comments on each.

For convenience, the tools open to lawmakers are grouped in ascending order of intrusiveness. **Option 0** (no change) is self-explanatory. **Options 1** and **2** would let courts decide which databases should be protected on a case-by-case basis. These Options would probably be most useful in a world where some (but not all) databases were inadequately protected against free riders. Finally, **Options 3** through **7** would grant protection to all database providers. These Options would be most appropriate in a world where virtually all databases faced significant threats of free ridership.

Option 0: No Change in Existing Law. Part I has shown that many providers are willing to offer databases based on self-help alone. Furthermore, Part II.A has shown that most scientific databases display sufficient creativity to qualify for copyright protection even after *Feist*. This probably affords a modest amount of protection despite the fact that competitors are still free to copy data if they rearrange it "creatively."

In the end, it is an empirical question whether existing protection strategies (i) prevent suppliers from investing in new databases, or (ii) unacceptably distort database content. Perhaps the most that can be said is that the Committee should seek evidence of need.

Option 1: Judge-Made Unfair Competition Law. This is the traditional formulation which governed database protection in the US prior to *Feist*. The most recent and sophisticated statement of the doctrine is found in the Second Circuit's *NBA* decision, which asks courts to consider five separate "elements" before deciding whether protection is appropriate:

1. Whether the database owner generates or collects information at some cost or expense;

2. Whether the value of the information is time sensitive;

3. Whether the defendant's use of the information constitutes free-riding on the plaintiff's costly efforts to generate or collect it;

4. Whether the defendant's use of the information is in direct competition with a product or service offered by the plaintiff; and

5. Whether the ability of other parties to free-ride on the efforts of the plaintiff would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened.[90]

As applied by the Second Circuit, a failure to prove any one of these elements would be fatal to a database owner's claim for protection.

As previously noted, the Second Circuit's discussion in the *NBA* case was narrowly limited to so-called "hot news" cases by Element (2) (time-sensitive information). At the same time, it is interesting to ask what would happen if Congress or the Courts were to overrule or liberally expand this element.[91] In that case, it seems clear that Elements (1), (3), and (5) (which collectively encapsulate the usual arguments against free ridership) would authorize courts to extend protection on a case-by-case basis.[*] Such case-by-case flexibility would be particularly appealing if the Committee believed that some (but not all) databases were vulnerable to "free ridership."

In the past, Legislatures have often relied on judicial discretion to implement policy on a case-by-case basis. Nevertheless, sending database protection back to the courts has drawbacks. Because judicial elaboration takes time, it might be many years before would-be copiers received a clear understanding of when the defense could and could not be raised. For example, the seventy year hiatus between *INS* and *Feist* produced very little guidance.

Option 2: "Improved" Unfair Competition Law. This is the first of two alternative reform measures advocated by **Reichman & Samuelson**. Drawing on *INS* and its progeny, courts would use the following eight factors to determine whether "unfair extraction" had occurred:

1. the quantum of data appropriated by the user;

2. the nature of the data appropriated;

3. the purpose for which the user appropriated the data;

4. the degree of investment initially required to bring that data into being;

5. the degree of dependence or independence of the user's own development effort and the substantiality of the user's own investment in these efforts;

---

[*] Element (4)'s limitation to use of information "in direct competition with a product or service offered by the plaintiff" is more suspect. From an economic perspective, society wants investment incentives to reflect the potential value of a proposed database to *all* markets -- not just the ones that the owner happens to be in at any given time. Suzanne Scotchmer, "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 29 - 41 (Winter, 1991).

6. the degree of similarity between the contents of the database and a product developed by the user (even if only privately consumed);

7. the proximity or remoteness of the markets in which the database owner and user are operating; and

8. how quickly the user was able to come into the market with his or her product as compared with the time required to develop the original database.[92]

Reichman & Samuelson correctly note that courts could use the foregoing factors to identify instances in which "database suppliers are sometimes less vulnerable to free-riding injury than appears from superficial claims for relief."[93]

At the same time, Reichman & Samuelson's eight factors are just that -- "factors," not "rules." Even more than the Second Circuit, their schema would commit the task of developing bright-line rules to future judges.

Option 3: *Sui Generis* Protection for a Limited Term. Case-by-case approaches make little sense if almost all databases are vulnerable to copying. Under such circumstances, the "exclusive rights model" found in copyright and patent law is probably appropriate.

This still leaves the question of how much protection is needed. One way to adjust this parameter is to provide protection for a fixed number of years. Presumably, the periods chosen should be related to the typical time that database providers need to recoup their investments. However, there are at least two types of investments:

Case (i): Creation of the underlying database,

and

Case (ii): Updating and maintenance

In general, policymakers might decide that only one of these cases actually needs protection. In a world dominated by Case (ii) investments, the long protection periods associated with Case (i) are probably unnecessary. Even if Case (i) protections are needed, moreover, very few US firms willingly make investments that cannot be recouped in 15 to 25 years. For this reason, the time periods found in the EU Directive, WIPO draft treaty, and HR 2652 are almost certainly too long.

Finally, the question of whether to protect "updates" is separate and distinct from that of Case (i) protection. If Case (ii) protection is granted, it probably should not last significantly longer than the mean time between updates -- *i.e.*, one or two years.*

Option 4: *Sui Generis* Protection With Non-Profit/Academic Exemptions. A second way to adjust the amount of protection afforded by a *sui generis* statute is to permit copying in certain clearly defined circumstances. One added advantage of this approach is that exemptions can be tailored to protect socially useful activities. Database providers who want to see their products disseminated on a non-profit basis should be encouraged. Exemptions can shelter non-profit databases against inadvertent statutory pressures to privatize (**Issue 6**).

The EU Directive allows member states to exempt copying "for the purposes of illustration for teaching or scientific research" or a "non-commercial purpose." An earlier NRC panel has similarly recommended that any future legislation should embrace the principle that "Database owners should never possess the right to preclude access to otherwise publicly available data when sought for purposes of basic scientific research."[94]

Practically all of today's science and technology fields depend on one or more non-profit databases. For this reason, the effects of privatization are likely to be pervasive. Significant adverse impacts could include (i) reduced real government funding levels, and (ii) damage to the existing culture of science (**Issue 6**). Future legislation should move cautiously in this area.

Option 5: *Sui Generis* Protection With a Defense for Improved Databases. Since advocates of extended database protection usually base their arguments on free ridership, it might make sense to exempt copiers who are willing to incur substantial costs. This view fits naturally with the existing world, in which databases are typically created by combining, improving, and extending earlier products.

The principal drawback of such a defense is that "substantial improvement" is hard to define and would almost certainly require judicial elaboration. The defense would presumably be available to any copier who invested in improvements, updates, and/or extensions at levels comparable to those of the original owner. Short of this, there is no obvious way to determine how substantial the copier's improvements would have to be. The concept would probably require judicial elaboration over time.

---

* Various commentators have suggested that initial "start up" protection should be extended each time a database is updated. If the database has only been "updated," it makes little sense to extend "startup" protection a second time.

Option 6: "Shrinkwrap" Contract Reforms. Everyday experience suggests that the lawyers who write "shrinkwrap" and "clickwrap" contracts will continue to claim as many rights as possible -- even when those rights happen to exceed the normal scope of copyright. The only real question, therefore, is what the courts will enforce. The draft UCC provisions discussed above provide little guidance.

The unpredictability and uncertainty of asking the courts to evolve common law solutions to the database problem have already been discussed under **Option 1**. However, common law unfair competition is at least based on free-ridership and other relevant concepts. In contrast, the shrinkwrap doctrine tends to be more concerned with contract law concepts like "offer," "consent," and "unconscienability." Since these concepts have little or nothing to do with free ridership, reliance on the shrinkwrap doctrine is likely to divert attention from the public policy issues most relevant to databases.

Option 7: Administrative Solutions. In their preferred (second) solution, **Reichman & Samuelson** argue that all databases should be protected by automatic licensing according to a predetermined fee schedule.* Although they recognize that automatic licensing schemes have met with mixed reviews in the past, Reichman & Samuelson believe that these criticisms could be ameliorated by (i) using an industry-based "collection society" to set "baseline" license fees, and (ii) allowing would-be licensees to opt out of the baseline by negotiating fee schedules directly with the database's owner.[95]

Reichman & Samuelson are right to point out that the collection society concept has a history of mixed reviews. Potential problems include the following:

- **Need for Market-Based Solutions.** Economists have traditionally justified intellectual property because it creates a mechanism for turning private knowledge of R&D opportunities into socially optimal levels of investment. Reichman & Samuelson's proposal would replace this market mechanism with a collection society's judgment of what fees should be. For a particular database, the regulated price will either be lower than that required to cover costs (thereby jeopardizing investment) or higher (thereby deterring use).

- **Transaction Costs.** Allowing participants to contract around the collection society may reduce transaction costs, but will not eliminate them. (This is true for the same reason that allowing litigants to settle lawsuits has not put the court system out of business).

---

* Reichman & Samuelson also suggest using an initial "blocking period" in which no databases could be copied. *Id.* at 145-46. This is conceptually identical to *sui generis* protection (**Option 3**) and will not be discussed further.

- **Antitrust Concerns.** Reichman & Samuelson correctly note that their proposal could only be enacted after removing "any antitrust barrier that stands in the way . . ."[96] However, the dangers of collusion should not be minimized. Giving an industry-based collection society the power to set database prices would create a political lightning rod. If suppliers (consumers) eventually became dominant, the temptation to impose monopoly (monopsony) solutions could become irresistible.

Given these concerns, Reichman & Samuelson's proposal should be viewed with caution absent strong evidence that the existence of "niche markets" has created a "natural monopoly" requiring regulation. Even then, the issue of whether license fees should be set by an industry-based collection society remains an open question.

Conclusion

The principal argument for statutory protection is that firms do not create enough databases because this would require a large "up front" cost that is not currently protected. However, this report has found little evidence that lack of statutory protection has prevented the creation of new products. The Committee should ask witnesses for concrete examples where this has happened. The Committee should also ask whether the assumption of large "up front" costs is realistic. Most of the database industry's products may instead consist of updates and improvements whose cost can be recouped within a year or so.

This report *has* found evidence that "self-help" can cause distortions. From the vendor's perspective, these include overinvestment in updates, graphics, software, and other "enhancements" at the expense of databases themselves. From the consumer's perspective, self-help can unnecessarily restrict access to data. The Committee will have to decide how serious such distortions are and whether they constitute an adequate case for reform.

From a legal standpoint, the Committee should remember that virtually all commercially valuable data can be described as "compilations of information" and hence "a database." So-called *sui generis* protection is therefore unlikely to stay confined to a particular type of information for very long. Sooner or later, most commercially valuable information will probably end up receiving "database" protection. This may or may not be a sensible result, but that is the choice.

Finally, the benefits of reform must be weighed against likely costs. Potential problems include, but are not limited to, deterring the creation of new databases from earlier products; creating monopoly power within "niche" markets; making databases unaffordable by the same university researchers whose work typically advances knowledge in the first place; and damaging the culture of science through inappropriate privatizing and hoarding of information.

Throughout this century, most arguments for and against database protection have proceeded from relatively simple assumptions about why databases are created and how they are sold. This report has shown that the reality is much more subtle. The pending Workshop represents a unique opportunity to deepen and extend this understanding.

Endnotes

1. Laura D' Andrea Tyson and Edward Sherry, *Statutory Protection for Databases: Economic & Public Policy Issues* (1997) (report commissioned by the Information Industry Association). The breadth of this definition is intentional. Indeed, the European Union's Directive on Databases expressly extends to

> ". . . literary, artistic, musical or other collections of works or collections of other material such as texts, sounds, images, numbers, facts, and data [as well as] collections of independent works, data or other materials which are systematically or methodically arranged and can be individually accessed."

Directive at ¶ 17. In fairness, the Directive does include an *ad hoc* exclusion for "audiovisual, cinematographic, literary, or musical work as such." *Id.*

2. The *Gale Directory of Databases* describes itself as "easily . . . the most complete guide to the electronic database industry worldwide." Kathleen Lopez Nolan (ed.), *Gale Directory of Databases* (New York & London, 1995) *at* vi. Readers seeking an entry-by-entry description of the sample can find it at Exhibit 1. A particularly useful feature of the *Gale Directory* is Prof. Martha E. Williams' annual profiles of the industry.

3. Joel S. White (personnel communication).

4. Joel S. White (personal communication). Info-Trac copies the articles at Bay Area libraries.

5. This section is taken from a five hour interview between the author and Dr. Richard Firestone, the head of LBL's *Table of Isotopes* project. Curious readers will find full details in a memorandum attached as Exhibit 3. An earlier workshop has studied Brookhaven's related but distinct ENSDF database. *See,* National Research Council, *Bits of Power: Issues in Global Access to Scientific Data* (n.a.:1997) ("*Bits of Power"*) at Appendix C. A copy of Appendix C is reproduced at Exhibit 4 to this report.

6. This section is taken from a brief interview with Karen Hunter, who handles copyright issues and strategic planning for Elsevier's scientific journals and databases. Curious readers will find full details in a memorandum attached as Exhibit 5. Exhibit 5 also contains additional information reproduced from Reed-Elsevier's Web site.

7. J.H. Reichman, personal communication.

8. This section is taken from a four hour interview between the author and Thomas R. Slezak, head of bioinformatics for Lawrence Livermore National Laboratories human genome sequencing group. Full details can be found in a memorandum attached as Exhibit 6. A supplementary discussion of genome databases can be found in Appendix C to the NRC's *Bits of Power* study and is reproduced here as Exhibit 4.

9. In 1997, *Science* described the bottleneck this way:

> "Because the world's major biological databases are constructed differently, it is virtually impossible to devise search programs to tap into them all effectively. A user has to hop from one to the other using each database's search engine to retrieve information that comes in a variety of different formats."

The article went on to describe how a "group of leading pharmaceutical companies" was putting their "considerable weight behind the development of common standards." Nigel Williams, "Drug Firms Back Move to Link Databases," *Science,* Aug. 15, 1997.

10. *See, e.g.,* Nigel Williams, "Unique Protein Database Imperiled," *Science* May 17, 1996 (international reaction to threatened closure of SWISS-PROT database); Howard M. Cann, "After the Genome Database," *Science* March 13, 1998 (user comment on closure of "GDB" database). Dr. Cann's letter is particularly illuminating for its discussion of the current system and how it might be fixed:

> "In the post-GDB-project world, the user may have to click more often to find mapping information [at other Web sites] and perform interpretation and editing personally. Problems that might be expected in the absence of GDB coordination include recognizing duplicates of new markers and conflicting map locations from different resources.
>
> "Perhaps the community will get by with the available final copy of the GDB and with database 'shopping' on the Internet. If not, the international community may have to pull together to arrive at a solution. For instance, database host institutions could form a consortium for the purpose of reviewing new data and maps in a coordinated fashion before release to the public. External expert reviewers might volunteer efforts (similar to those of the 'editor' group of scientists that now review and edit GDB data) within the framework of such a consortium, injecting further assurances of quality and coordination. This type of program or something with a similar intent could be provided at a minimal cost increase and would continue to support the efforts of many scientists involved in mapping and eventually identifying genes underlying complex disorders."

Id.

11. "Incyte Serves Up Information, Part I," *In Vivo*, May 1996.

12. *See, e.g.,* Jon Cohen, "The Genomics Gamble," *Science*, Feb. 7, 1997. The database user exercised its right to terminate the partnership in late 1996. *Id.*

13. "Genetic Warfare," *The Economist*, May 16, 1998.

14. "Incyte Serves Up Information, Part II," *In Vivo*, May 1996.

15. Incyte's bioinformatics capabilities are summarized on its Web site. The interested reader can find selected Web pages included as part of Exhibit 6 to this report.

16. "Perkin-Elmer's Pharmacogenomics Spin-Off Creating a New Customer for Instrumentation," *Bioventure View,* June 1, 1998.

17. Except as noted, all information reported in Examples 6(a) through (f) is taken from Tyson & Sherry at pp. 3 - 6. Supplemental research has been taken from the Web and is collected at Exhibit 2. Examples 8 (g) through (h) are based on descriptions found in Appendix C to *Bits of Power* at pp. 209-10 (materials science), 210-212 (chemistry), 214- 216 (geophysics), and 217-218 (meteorology). A copy of Appendix C is attached as Exhibit 4 to this report.

18. Interview with Richard Firestone (Exhibit 3).

19. Interview with Karen Hunter (Exhibit 5).

20. *See, e.g., ProCD, Inc. v. Zeidenberg*, 86 F.3d 1447 (7th Cir. 1996) ("contracts about trade secrets may be enforced").

21. Although technologically less secure, CD-ROM makers often use the parallel strategy of encryption to block access to their databases. This type of self-help recently received a legal boost when the US Congress enacted P.L. 105-304 ("The Digital Millennium Act"). The statute establishes criminal fines and penalties for anyone who tries to defeat an electronic encryption system.

22. If anything, the statistic errs on the side of conservatism since it ignores products that advertise "irregular" updates.

23. *International News Service v. Associated Press*, 248 US 215 (1918).

24. *Id.* at 236.

25. *Id.* at 241.

26. Jack E. Brown, "Obscenity, Anonymity, and Database Protection: Emerging Internet Issues," *The Computer Lawyer* (Oct. 1997) (citations omitted). In 1942, a federal judge argued that *INS* would have been decided differently if it had been heard it that year. *Id.* at fn. 78.

27. *Feist Publications, Inc. v. Rural Telephone Service Co., Inc.*, 499 US 340 (1991).

28. *Id.* at pp. 342 - 45.

29. *Id.* at p. 344.

30. *Id.* at p. 348.

31. *Id.* at 352 - 53.

32. *Id.* at 354.

33. *Warren Publishing, Inc. v. Microdos Data Corp.,* 52 F.3d 950 (11th Cir. 1995).

34. *Key Publications, Inc. v. Chinatown Today Publishing Enterprises, Inc*., 945 F.2d 509, 514 (2d Cir. 1991).

35. *Bellsouth Advertising & Publishing Corp. v. Donnelly Information Publishing, Inc.* 999 F.2d 1436, 1441 (11th Cir. 1993).

36. *Id.* at 1441.

37. *Mason v. Montgomery Data, Inc.,* 967 F.2d 135, 139 (5th Cir. 1992).

38. *Nester's Map & Guide Corp. v. Hagstrom Map Co.,* 796 F.Supp. 729, 733-34 (E.D.N.Y. 1992).

39. *CCC Information Services, Inc. v. MacLean Hunter Market Reports, Inc.,* 44 F.3d 61, 67 (2d Cir. 1994). According to *CCC,* an author's "loose judgment" that "vast regions" of the US could be treated as a single market was also protectable.

40. *Warren Publishing, supra,* at 951 - 52.

41. *Skinder-Strauss Associates v. Massachusetts Continuing Legal Ed., Inc.,* 914 F.Supp. 665, 675 (D. Mass. 1995).

42. *Cable News Network, Inc. v. Video Monitoring Services of America, Inc.,* 940 F.2d 1471 (11th Cir. 1991) at 1485. The opinion was subsequently vacated on other grounds and is cited here as an indication of what future courts might decide if faced with the same question. *Cable News Network, Inc. v. Vide*o *Monitoring Services of America, Inc.,* 949 F.2d 378 (1991).

43. *National Basketball Assn. v. Motorola, Inc.,* 105 F.3d 841 (2d Cir. 1997).

44. *See* Interview with Karen Hunter (Exhibit 5 hereto).

45. *Sinai v. California Bureau of Automotive Repair,* 25 USPQ 2d 1809, 1811 (N.D. Cal. 1992).

46. *ProCD, Inc., supra*. One particularly noteworthy aspect of the *ProCD* decision was the court's statement that it would "refrain from adopting a rule that anything with the label 'contract' is necessarily outside the preemption clause." *Id.* at 1455.

47. *Vault Corp. v. Quaid Software Ltd.,* 847 F.2d 255 (5th Cir. 1988).

48. Council Directive No. 96/9/EC, *O.J.*L 77/20 (1996). The Directive itself is not intended to be legislation. Instead, it sets forth requirements that member states must satisfy by enacting "at least materially equivalent" statutes. *Id.* at ¶ 32. A copy of the Council's Directive is attached as Exhibit 7 to this report.

49. *Id.* at Art. 1, ¶ 2.

50. *Id.* at Art. 5, subpart (a).

51. *Id.* at Art. 5, subpart (b).

52. *Id.* at Art. 5, subparts (c) -(e).

53. *Id.* at Art. 7, ¶ 1.

54. *Id.* at Art. 10, ¶ 2.

55. *Id.* at Art. 10, ¶ 3.

56. *Id.* at ¶ 56.

57. *Id.* at Art. 6, ¶ 2(b).

58. *Id.* at Art. 6, ¶ 2(d).

59. 17 USC § 107.

60. "Basic Proposal for the Substantive Provisions of the Treaty on Intellectual Property in Respect of Databases to be Considered by the Diplomatic Conference" dated August 30, 1996 (hereinafter "*WIPO*"). Interested readers will find a copy of the *WIPO* draft at Exhibit 8 to this report.

61. *Id.* at Art. 2, ¶ (i). The definition would have specifically included "collections of literary, musical or audiovisual works or any other kind of works, or collections of other materials such as texts, sounds, images, numbers, facts, or data representing any other matter or substance. It is worth pointing out that in addition to many kinds of works and other information materials, databases may contain collections of expressions of folklore." *Id.* at comment 2.02.

62. *Id.* at Art. 2, ¶ (ii) and Art. 3, ¶ (1). The definition of "substantial part" was further amplified in a note:

> "The substantiality of any portion of the database is assessed against the value of the database. This assessment should evaluate the qualitative and quantitative aspects of the portion, although neither aspect is more important than the other . . . The value of a database refers to its commercial value. This value consists on the one hand of direct investments made in the database and on the other hand of the

> expected market value of the database. This assessment may also take into account diminution of market value that may result from the use of the portion, including the added risk that the investment in the database will not be recoverable. It may even include an assessment of whether a new product using the portion could serve as a commercial substitute for the original, diminishing the market for the original."

Id. at Note 2.09. The concept of an "investment" included any and all "human, financial, technical or other resources" devoted to "the collection, assembly, verification, organization, or presentation of the contents of the database." *Id.* at Note 2.10 (iv).

63. *Id.* at Art. 5, ¶ (1). The accompanying notes emphasized the point by explaining that such exceptions "may never conflict with normal exploitation of the database" and could not "unreasonably impair or prejudice the legitimate interests, including economic interests, of the rightholder." *Id.* at Note 5.01.

64. "*WIPO"* at Art. 8.

65. Jocelyn Kaiser (ed.), "Treaty on Database Access Stalled," *Science,* Dec. 20, 1996.

66. The Digital Millennium Copyright Act of 1998 was subsequently enacted as Public Law 105-304.

67. HR 2652 at § 1201.

68. *Id.* at § 1202.

69. *Id.* at § 1203(a).

70. *Id.* at § 1202(b) - (c).

71. *Id.* at § 1203(d) (emphasis supplied). The reference to "potential markets" would have been more restrictive than the corresponding EU Directive, which permits copying "for the purposes of . . . scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved." Directive at Art 9, subpart (a). The "potential markets" language was dropped shortly before the bill went to conference committee. Paul Uhlir (personal communication).

72. *Id.* at § 1203(e).

73. *Id.* at § 1204.

74. *Id.* at § 1206(d). Courts would also have been given discretion to reduce damages for any employee of a nonprofit educational, scientific, or research institutions who "believed and had reasonable grounds for believing that his or her conduct was permissible under this chapter." *Id.* at 1206(e).

75. *Id.* at § 1207.

76. UCC 2B-110 (August, 1998 draft). A copy of the draft provision with accompanying notes can be found at Exhibit 10 to this report. Interested readers can view the entire file at <http://www.law.upenn.edu/library/ulc/ucc2b/2b898.htm>.

77. *Id.* at Note 2.

78. *Id.* at Note 3. Significantly, the Reporter adds that state court judges "may look to federal copyright and patent laws for guidance on what types of limitations . . . ordinarily seem appropriate." *Id.* This suggests that federal law may provide persuasive reasons why state courts *should* refuse to enforce particular licenses even where it does not directly command them to do so.

79. *Id.* at Note 1.

80. *Id.* at Note 3.

81. *Cf.*, Reporter's Note 3 to draft UCC provision 2-105 (copyright statute permits "contractual restrictions on use.")

82. *Id.* at Note 3 (emphasis supplied).

83. J.H. Reichman & Pamela Samuelson, "Intellectual Property Rights in Data?," *Vanderbilt Law Review* Vol. 50, p. 51 (January 1997).

84. Walter Nicholson, *Microeconomic Theory: Basic Principles and Extensions* (6th ed. 1995) at pp. 625 - 28.

85. Michael Heller & Rebecca Eisenberg, "Can Patents Deter Innovation? The Anticommons in Biomedical Research," *Science* (May 1, 1998) 280: 698 - 701; *see also,* Suzanne Scotchmer, "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 29-41 (Winter, 1991).

86. Jerry Green & Suzanne Scotchmer, "On the Division of Profit Between Sequential Innovators," *Rand Journal of Economics*, 27: 322-31 (1996).

87. US Patent and Trademark Office, *Report on and Recommendations from April 1998 Conference on Database Protection and Access Rules* (July, 1998) at p. 16.

88. *Id.* at pp. 14 - 17; Terry M. Sanks, "Database Protection: National and International Attempts to Provide Legal Protection for Databases," *Florida State University Law Review* 25:992 (1998).

89. *Directive* at ¶ 11. ("Whereas there is at present a very great imbalance in the level of investment in the database sector . . . between the Community and the world's largest database producing third countries.") At first blush, the EU's logic seems paradoxical since greater incentives would also encourage US companies to compete even harder. However, the EU may believe that American companies have *already* decided to enter the database market. If so, additional protection might persuade risk-averse European firms to enter the market without eliciting still more investment by the Americans.

90. *National Basketball Assn., supra,* at 852 - 53.

91. *See,* US Patent and Trademark Office *Report, supra* at p. 6 (reporting suggestion by Professors Ginsburg & Reichman).

92. Reichman & Samuelson, *supra,* at pp. 142-43.

93. *Id.* at 143 & fn. 423.

94. *Bits of Power, supra,* at 166.

95. Reichman & Samuelson, *supra,* at 146 - 50.

96. *Id.* p. 148.