

A PRIMER ON WIPO

by James Love

Database Extraction Rights*

The World Intellectual Property Organization (WIPO) considered in December 1996 a new treaty that would have required most countries (including the United States) to severely curtail the public's rights to use public domain materials stored in "databases." Some experts say it could be the "least balanced and most potentially anti-competitive intellectual property rights ever created." The database treaty was being pushed by large publishing companies in response to the 1991 U.S. Supreme Court decision in *Feist Publications, Inc. v. Rural Telephone Service*. In *Feist*, the Court rejected a claim of copyright for data from a telephone directory's white pages, saying that facts cannot be copyrighted, and that obvious items such as listing names, addresses, and telephone numbers in alphabetical order, are not sufficiently creative to qualify for copyright protection. The decision rejected the "sweat of the brow" theory of copyright.

Compilations of data or documents, including materials from the public domain, can receive protection under copyright if the creator of the compilation can show originality in the selection and arrangement of the data. Comprehensive databases, which can be expensive to create, confront problems under copyright laws because (almost by definition) they are not original in terms of the selection of the materials.

Electronic database publishers have sought to protect their data through contracts with their customers. These contracts often place restrictive conditions on the reuse or dissemination of the data. (See Pam Samuelson, "Legally Speaking: Legal Protection For Database Contents." 39 *Communications of the ACM*, Nov. 1996, for a discussion about this approach.)

* This two part article and accompanying case study were originally published on *Info-Policy Notes*, an electronic journal. The database issue was subsequently tabled, but the questions involved have not been resolved and are still extremely important —

In other cases, database vendors permit online searching, but do not distribute the complete database itself.

Publishers are looking for stronger protection, and are lobbying hard to obtain a new *sui generis* property right to protect the contents of databases. The publishers' first success was the adoption of a controversial proposal for database extraction rights in the European Union (EU), and by gaining the support of the Clinton Administration and the EU to propose a very similar measure as an amendment to the Berne Copyright Convention. Legislation to implement this form of data use regulation was introduced in the 104th Congress [HR 3531], but there were no hearings on the measure, and in Nov. 1996 the Clinton Administration did not support the language of HR 3531.

Despite the controversial and far reaching nature of the database protection proposal and the lack of discussion on its impact in the United States, the Clinton Administration asked for quick approval of the database treaty at a December 1996 meeting in Geneva hosted by the World Intellectual Property Organization (WIPO). The main Administration advocate in support of the publishers' position was Bruce Lehman, Chair of the Patent and Trademark Office (PTO), a person widely considered an intellectual property rights zealot.

The Complexities and Dangers in Creating a New Property Right for Data

While many people are sympathetic to the general idea of a "sui generis" form of protection for databases, there is enormous concern about the complexities of creating a new property right that has the potential to create private monopolies on data and documents that have traditionally been in the public domain. It is often said that "the devil is in the details," and this is certainly true for the database protection proposal. A handful of database vendors have quietly crafted a proposed treaty and law that creates a nightmare for researchers and value-added publishers. In discussing the development of the EU database proposal, J. H. Reichman and Pamela Samuelson say that "lobbying pressures converted the final version into one of the least balanced and most potentially anti-competitive intellectual property rights ever created" (in "Intellectual Property Rights In Data: An Assault On The Worldwide Public Interest In Research And Development," in the *Vanderbilt Law Review*, 50).

The database vendors have sought to vastly expand the ability of database owners to regulate and restrict the public's rights to use data, without the types

of safeguards which exist in copyright law today. In this respect, it is important to understand that as a *sui generis* property right, the database extraction rights are not part of the of the copyright regime, and the entire doctrine of fair use of data will not apply to data protected under the proposed database extraction rights treaty and legislation. Moreover, under the WIPO proposal these new data rights would be retroactive, affecting countless databases already in existence.

Digression on West Publishing and the Definition of a Database

The Feist decision was particularly troubling for West Publishing, a company that wants to maintain its monopoly on the citations and corrected text for many court decisions. West is the only comprehensive publisher of federal circuit and district court opinions and state court opinions from all 50 states. The page numbers of the West court reporters are the basis for authoritative citations used by scholars and lawyers. As a reporter of decisions, West also makes corrections to the text of court opinions, typically after working with the judge who wrote the opinion. West wants to prevent others from using their page numbers or the corrected text of court opinions, and it is often in court trying to prevent its would be competitors from doing so.

West is now involved in at least two law suits over its assertions of copyright of the page numbers, and one law suit over the issue of the copyright to the text of the corrected court opinions. Most copyright experts think that West will lose its court case on the issue of its page numbers, and West will also be hard pressed to claim it can copyright the corrections to the text of court opinions—particularly for the US federal courts, since U.S. copyright laws exclude the works of federal employees.

Most people think that the corrected text of court opinions, and the citations to those opinions, should be in the public domain, and that the West monopoly has delayed the development of new information products and services for legal researchers. No one seriously argues that the court opinions would not be published without a West monopoly. West is among the private sector publishers who have successfully lobbied the EU and the Clinton Administration to extend database protection proposals to print products by defining a database so broadly that it will include any collection of facts, data, or documents regardless of the media. If the database protection proposals are enacted, West would have a firm monopoly on decades of judicial citations and

corrections to judicial opinions.

What is a Database? What Isn't a Database?

The treaty would protect “any database that represents a substantial investment in the collection, assembly, verification, organization or presentation of the contents of the database.” This term should be understood

to include 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 [and] may contain collections of expressions of folklore. [The] protection shall be granted to databases irrespective of the form or medium in which they are embodied. Protection extends to databases in both electronic and non-electronic form [and] embraces all forms of media now known or later developed....Protection shall be granted to databases regardless of whether they are made available to the public. This means that databases that are made generally available to the public, commercially or otherwise, as well as databases that remain within the exclusive possession and control of their developers enjoy protection on the same footing.

In other words, a lot of water will go under this bridge.

What Are Extraction and Utilization Rights?

“The maker of a database eligible for protection under this Treaty shall have the right to authorize or prohibit the extraction or utilization of its contents.” What is “extraction”? Extraction is defined as, “the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form.” “Extraction...is a synonym for ‘copying’ or ‘reproduction’...by ‘any means’ or ‘any form’ that is now known or later developed.”

“Utilization” is defined as “making available to the public all or a substantial part of the contents of a database by any means, including by the distribution of copies, by renting, or by on-line or other forms of transmission,” including the right to control the use of the data “at a time individually chosen by each member of the public.”

What is a "Substantial Part" of the Database?

The treaty set out tests for determining if an extraction is "substantial," and these tests are both highly anti-competitive, and extremely broad in scope. The "substantiality" of a portion of the database is assessed against the "value of the database," and considers "qualitative and quantitative aspects," noting that "neither aspect is more important than the other... This assessment may also take into account the diminution in 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." Then the treaty added that a "substantial part" means any portion of the database, "including an accumulation of small portions... In practice, repeated or systematic use of small portions of the contents of a database may have the same effect as extraction or utilization of a large, or substantial, part of the contents of the database."

To implement this legislation in the US, the only types of data use that would not be regulated would be "insubstantial" parts, "whose extraction, use or reuse does not diminish the value of the database, conflict with a normal exploitation of the database or adversely affect the actual or potential market for the database." Under this language, a database owner could say that it might in the future want to charge for each transmission of a fact or an element of a database as part of its "normal exploitation" of the database. With the Internet and digital cash this claim is likely to be made. The public would not have "fair use" rights, since fair use is only defined in matters involving copyright.

How Long? 15 Years, 25 Years, or Forever?

The treaty would have required a minimum term of protection (15 years in the EU proposal, and 25 in the United States proposal) for the database. But this would be extended each time the database is revised or enhanced. According to the draft treaty, "any substantial change to the database, evaluated qualitatively or quantitatively, including any substantial change resulting from the accumulation of successive additions, deletions, verifications, modifications in organization or presentation, or other alterations, which constitute a new substantial investment, shall qualify the database resulting from such investment for its own term of protection."

The provision on revisions raises the specter that protection for many databases will be perpetual. This could indeed be the case if the original versions of the database are only "licensed" by the vendor for a limited period of time, so that the only available versions would be the new ones, which would have a new term of protection. (Database vendors write these restricted use licenses now.)

What About Government Information?

Much of the lobbying for the *sui generis* database proposal was designed to enable database vendors to protect collections of government documents. The treaty would permit countries to have special rules for "databases made by governmental entities or their agents or employees." However, this exemption will not include cases such as the West Publishing reporting of court decisions, where West is acting as an unofficial agent for the courts. In the US enabling legislation, protection is not given to a database made by a governmental entity, but protection could not be excluded from companies if a database's "contents have been obtained from a governmental entity." There is no provision to exempt databases created by private parties, like West, LEXIS, and literally thousands of other firms, when they act as contractors to government agencies. For example, West is a contractor for some courts in receiving electronic filing of briefs. Under the U.S. legislation, the database of briefs collected by West for the Courts would be protected. Likewise, the Security and Exchange Commission's EDGAR public disclosure filings which are managed by LEXIS would be covered.

The Clinton Administration has gone to court in at least two cases to avoid releasing documents under the Freedom of Information Act (FOIA) when West Publishing has asserted intellectual property rights claims to elements of the data. In the FLITE case, the Clinton administration successfully argued that it did not have to release US Court opinions collected by the Air Force at public expense that contained West "corrections" and enhancements. It appears as though government entities will be permitted to avoid FOIA completely if they use private contractors, and write contracts which permit agency access to data (extraction), but do not permit disclosure to the public.

What About Fair Use Rights?

As noted several times, the public has rights, often taken for granted, under the copyright "fair use" doctrine. This includes commercial and non-commercial

fair use. The fair use rules involve public interest balancing tests. The *sui generis* database proposal doesn't include or incorporate public fair use rights. It is difficult to know how this will play out in practice. Under the treaty language, governments "may, in their national legislation, provide exceptions to or limitations of the rights provided in this Treaty in certain special cases that do not conflict with the normal exploitation of the database and do not unreasonably prejudice the legitimate interests of the rightholder." The key terms here are "normal exploitation of the database," and "legitimate interests" of the rightholder.

In the US legislation, "a lawful user of a database made available to the public or placed in commercial use is not prohibited from extracting, using or reusing insubstantial parts of its contents, qualitatively or quantitatively, for any purposes whatsoever." But as noted earlier, the term "insubstantial" is constrained by the scope of the business opportunities that are perceived by the database vendor. Not only is "insubstantial" limited to those uses which do not diminish the value of the database, but insubstantial must also not "conflict with a normal exploitation" of the database, or adversely impact the "actual or potential" market of the database. Moreover, the "normal exploitation" of the database seems to be defined in such a way that the vendor can assert that a transmission of a database element on the Internet would be an infringement if the company has a mechanism or even aspirations to charge for the information, and the cumulative impact of many small transactions would diminish the value of that service. The treaty would require countries to provide protection prospectively for databases already on the market. Countries could exempt older databases from protection for up to two years.

Prohibitions on Technologies to Defeat Protection

Finally, as in the proposed Internet copyright treaty and legislation, the database extraction rights proposal is accompanied by very strict prohibitions against the "importation, manufacture or distribution of protection-defeating devices." This is defined as "any device, product or component incorporated into a device or product, the primary purpose or primary effect of which is to circumvent any process, treatment, mechanism or system that prevents or inhibits any of the acts covered by the rights under this Treaty." The US legislation contains similar provisions, plus a whole section which would make it a federal crime to interfere with "database management information." Persons would face up to 5 years in jail and a \$500,000 fine for doing such things as providing or

disseminating false database management information, or removing or altering any such information. It would seem that simply tearing the cover off a telephone book (a protected database under the treaty) would be a violation of this provision.

Government Proposes New Regulation of Sports Statistics and Other "Facts" — A Case Study of WIPO in Action

Sports fans in the United States will be surprised to learn that US Government officials have pressed for the adoption of an International treaty (the World Intellectual Property Organization treaty) that would (if enacted) significantly change the ways sports statistics are controlled and disseminated. The treaty isn't specifically directed at sports statistics — it is a much broader attempt to create a new property right in facts and other data now in the public domain — but it will have an enormous impact on the legal rights exercised by the National Football League (NFL), Major League Baseball (MLB), the National Basketball Association (NBA), the National Hockey League (NHL) and virtually all other professional or amateur athletic leagues. (The same treaty will radically affect the way that stock prices, weather data, train schedules, data from AIDS research and other facts are controlled, but this essay will focus on the issue of sports statistics, a topic that illustrates the broad impact of the treaty).

This comes at the same time the NBA and other sports franchises are stepping up their efforts to control the real-time dissemination of sports statistics through the Internet or with wireless paging devices. The treaty, however, addressed different and much more fundamental issues regarding ownership of information.

If the treaty were to be approved and implemented, sports leagues would have far broader powers to dictate the terms and conditions under which sport statistics are reported and disseminated. Nolan Ryan's Earned Run Average (ERA), the number of tackles or quarterback sacks by Lawrence Taylor, Cal Ripken's career batting average, Bobby Hull's career assists, the number of steals by your favorite NBA point guard, and similar information will be "owned" by sports leagues. According to the proposed treaty (and legislation introduced in the 104th Congress to implement the treaty), the NFL, NBA, NHL and MLB would have the right to prevent anyone from publishing these and other statistics without express permission from the sports league. This

includes the right to control access to the historical archives of sports statistics, and even to dictate who can publish the box scores from a game or print a pitcher's ERA on the back of a baseball card.

The treaty seeks, for the first time, to permit firms to "own" facts they gather, and to restrict and control the redissemination of those facts. The new property right would lie outside (and on top) of the copyright laws, and create an entirely new and untested form of regulation that would radically change the public's current rights to use and disseminate facts and statistics. American University Law Professor Peter Jaszi recently said the treaty represents "the end of the public domain."

Who Will Own Facts?

The supporters of the Treaty note that persons can independently collect data, and the US legislation says "nothing in this Act shall in any way restrict any person from independently collecting, assembling or compiling works, data or materials from sources other than a database subject to this Act." Unfortunately, this will only be helpful in those cases where there will be a separate non-protected source for the data or documents. If the entity which creates the initial data or documents qualifies for the database extraction right under the proposed treaty, the data itself will be monopolized. All sorts of data will be protected at the source under the database treaty, and may never enter the public domain. There are also the practical problems relating to the costs of data collection. The telephone companies obtain directory information when you become a subscriber, and it is practically impossible to independently collect this data. Databases of IP addressees collected by Network Solutions will be covered, giving Network Solutions broad new rights in how that data is utilized by ISPs.

Who Wants the Treaty?

In the 1991 Feist decision (see the previous article for a summary), the US Supreme Court ruled that, in general, "facts" could not be copyrighted by anyone. The Feist decision alarmed several large database vendors, who crafted this new property right that would protect facts, and just about everything else. (The vendors have already succeeded in obtaining a directive on the database proposal from the European Union, although no European country has yet

passed legislation to implement the treaty). The most active supporter of this new property right is West Publishing, the Canadian legal publisher (see the previous article). A West Publishing employee chairs a key American Bar Association subcommittee which wrote a favorable report on the treaty. A number of very large British and Dutch database vendors are also lobbying hard for the treaty. Other database vendors want to protect scientific data or other non-copyrighted government information they publish. In seeking to protect these items, the treaty was written to stamp "owned by" labels on a vast sea of information now in the public domain. (Copyright experts J.H. Reichman and Pamela Samuelson say it is the "least balanced and most potentially anti-competitive intellectual property rights ever created.")

There has been an active debate within the Clinton Administration over the proposed treaty. Bruce Lehman, the controversial head of the Patent and Trademark Office was pushing for adoption of the treaty last December. Most administration officials don't have a clue what the database treaty would do. Some people think it is a minor tinkering with the current copyright law. No one in the government sought to understand the significance of the proposal in terms of the new rights to "own" facts, and until recently no one was aware that the treaty was so broad that it would change the way sports or financial statistics were controlled.

What Does This Have to Do With Sports Statistics?

Since facts cannot be copyrighted according to the Feist decision, the supporters of the treaty have framed this as a new *sui generis* property right, which will have a separate statutory framework. "Originality" or "authorship" will not be required. "Texts, sounds, images, numbers, facts, or data representing any other matter or substance," will be protected. The information can be stored in "all forms or media now known or later developed." Both published and confidential information will be covered. The only thing required is a "substantial investment in the collection, assembly, verification, organization or presentation of the contents" of the protected work. The "rightholder" will have extremely broad powers to "authorize or prohibit the extraction or utilization" of the information from the protected database.

It takes a while for the implications of this new system to sink in. Some facts can be independently gathered, like the number of baseball games played in a year, the winners or losers of a tennis match, or the scores of a football game.

For these data, there may exist several sources for the data. However, other facts are, by their very nature, only available from a single source, and will be controlled by monopolies. For example, baseball leagues employ scorekeepers who determine if a batter is credited with a hit or if a fielder committed an error, if a hit is reported as a single or double, or if an errant pitch is scored as a wild pitch or a passed ball. The league makes a "substantial investment" in the collection and maintenance of this data, which it disseminates to the press, and also stores and maintains in a database, through an arrangement with the Elias Sports Bureau. These data cannot be independently collected — and under the proposed database treaty, the league would own the facts themselves, and could dictate the terms under which these facts are published or redisseminated.

The NFL employs four persons who keep track of the play-by-play action for each game. They write up four separate reports, which are used to create a single official "box score." The final product is supervised by the Elias Sports Bureau, as a "work for hire" product, which is owned by the NFL. The NFL box score is very detailed, and includes analysis of each play. It records the league's statistics for the number of yards gained (or lost) on each play, who is credited with a tackle or a quarterback sack, or the number return yards on a kickoff or pass interception, and many other items. While someone who attended a football game could make an independent estimate of these items, it would likely be different from the official statistics, due to the inherent difficulty in measuring or assigning credit for performance on the field. The NFL's box score is given to the press, which uses the data to create its own news media reports.

An attorney who represents the NFL told us that the NFL has an interest in ensuring that there is an "official" source of the statistics, which are gathered with an appropriate standard of care and that the NFL "protects the official designation" of its statistics. These data are used or making decisions on the Hall of Fame, and to create special reports and information products, which the NFL provides to third parties, often for a fee.

Virtually all of the major sports leagues have some system for creating statistics, disseminating the information to the press, storing the historical data, and marketing the statistics commercially. Major League Baseball, the NFL, and the NBA work with Elias, while the NHL has its own in-house system. There is little doubt that the process by which these statistics are generated will qualify for protection, under the treaty's minimal requirement that the league demonstrate it has made a "substantial investment in the collection, assembly, verification, organization or presentation of the contents" of the database. The

work-for-hire "media sheets," "box scores," and other press handouts which report the statistics would be considered database elements, and reporting of statistics from these products would be subject to an entirely new type of licensing and control by the leagues which is far stronger than that which exists under copyright law.

The leagues have various methods of selling their "official" branded statistics. There are also many competitors who build databases from a variety of sources, including the published boxscores that appear in daily newspapers, and probably the books and reports published by the leagues. The leagues do not currently assert "ownership" in the statistics directly, even as they try to prevent others from referring to the data as "official" statistics, but they are trying to prevent real time reporting of game statistics and situations over Internet or paging technologies.

The NBA told us that it permits accredited journalists to report scores from NBA games three times each quarter, and that it considers the minute to minute reports a "misappropriation" of its ability to sell performance rights for the event. The NFL takes a similar position with respect to its games. STATS, Inc. is a firm that provides real time scores and play-by-play descriptors to a variety of online and wireless information services. According to the NFL and the NBA, STATS, Inc. hires people to watch television broadcasts of the games, and type the play-by-play information into personal computers, which are linked to the STATS, Inc. computer network services. An example of this type of service that uses STATS, Inc. as a supplier of statistics is Instant Baseball. Disputes over the real time Internet broadcasts of game situations and scores could well end up before the U.S. Supreme Court, as a test of the First Amendment. In the NBA case involving Motorola, STATS, Inc. and America Online, the NFL and other leagues have filed amicus briefs in support of the NBA position, while the *New York Times* has filed a brief in support of Motorola, STATS, Inc. and AmericaOnline.

The NBA has also discovered HR 3531, a version of the database protection proposal that was introduced in the U.S. Congress last spring. The NBA is looking at HR 3531 to see if it would provide a legislative remedy for their dispute with Motorola, *et al.* At present, none of the leagues currently prevent anyone from publishing statistics after a game is over, because it is assumed that the statistics (facts) are in the public domain, once the broadcasts are over. But this would likely change if the database treaty is enacted. One league official told me, "no matter how appalled I am at this proposal personally, as a civil

libertarian, my client may have interests as a rightholder that it will want to exercise." A lawyer for the NFL said that the NFL might not want to do anything — he thought the free dissemination of statistics brought its own benefits, in terms of increased fan interest. But he also said, the treaty would allow NFL to "do quite a bit of stuff" in terms of new licensing arrangements or other ventures, if it wanted to.

As a *sui generis* property right, the database proposal does not incorporate the fair use principles from copyright that reporters and value-added publishers often take for granted. The leagues would be able to require licenses to publish box scores or other statistics in any media. One can imagine a world where the leagues wouldn't require licensing of box scores to print-based periodicals like daily newspapers, but that a much more controlled regimen would evolve on the Internet. The leagues could require licensing of box scores and other statistics for Internet publications, or linking to the leagues own web sites, such as www.nba.com, www.nfl.com, or www.nhl.com. The Internet is, after all, a very easy place to locate and police violations of intellectual property rights — through a simple AltaVista search. This would also likely lead to major changes in the market for baseball (and other sports) cards, which typically feature key statistics on the flip side of the card. The new database extraction rights would prohibit any unauthorized extraction or reuse of data that had economic value to the leagues.

Conclusion

Of course, this treaty dealt with a lot more than sports statistics. It would do the same thing for information on stock prices that is generated by a stock exchange. It will radically change the market for weather information. There is concern in Europe over the control of train schedules. Private schools could use the new data extraction right to prevent unauthorized publication of data about its students' test scores or postgraduate placement statistics (both generated from a database). The treaty would radically change the rights to use information from gene sequencing, or hospital cost-benefit studies. It will obviously do much much more.

Consider this: the treaty, which was designed to protect West Publishing's legal reporters, was written so broadly that it would have defined even the daily newspaper as a "database" element. Since the new property right is additive to all rights claimed under copyright, every publisher will claim the additional

protection, by saying each issue of the newspaper is a database element. (Virtually all newspapers today are archived in databases). The consequences of this are astounding, since every fact and article in every newspaper would have the new stronger form of protection, which would not include any public fair use rights.