Database Protection: Are Laws Threatening To Destroy Our Building Blocks Of Knowledge?

Julia Alpert Gladstone
Bryant College, Smithfield, RI, USA

gladsto@bryant.edu

Abstract
This paper examines the various regimes that are used to protect databases to suggest that the continued progress of science and technology that has enabled economic prosperity will be fostered by less regulation. The diversity between and within each of these regimes reflects fundamentally different views of intellectual property. Technology, specifically digitalization that has facilitated the creation, replication and easy dissemination of information has changed the value of information and threatens to create a striated society of information "haves" and "have-nots" due to enclosure mechanisms. As technology advances, the laws which we implement to build upon the existing intellectual property infrastructure must be developed with care to preserve the careful balance of the public good and private interest that has maintained the past 200 years of "progress of science and useful arts." The author suggests ways to structure a database to encourage or reward database developers while simultaneously fostering the advancement of science.

Introduction
Digitized information and the tools to use it represent approximately 6.6% of the world's gross domestic product and expenditures in this industry are expected to surpass $3 trillion by 2003 (Cohan, 2001). The importance of “databases” or “compilations of information” may be best illustrated with examples of how integrated into our lives databases have become. Whether one flips through a listing of TV channels, searches the Internet for a chocolate cake recipe, grabs a listing of today’s events in the city or reviews a journal for statistics pertinent to one’s career, one has relied upon a database. The importance of data to everyday life is self-evident. Its importance to the global development is best stated by Nobel laureate Joshua Lederder, “data are the building blocks of knowledge and the seeds of discovery. The assembled record of scientific data and resulting information is both a history of events in the natural world and a record of human accomplishment” (Reichman and Uhlir, 1999, p 793).

As we look towards the second session of the 107th session of Congress, there is strong support to pass more protective database legislation in response to Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases, 1996 O.J. (l77) (the E.U. Database Directive). Such legislation would supplement the existing intellectual property, misappropriation and contract laws that are already in place. The existing legal protections often work in conjunction with technological mechanisms or safeguards. Over the past five years the several database protection bills that have worked their way through various committees of both Houses incorporated provisions which would establish a sui generis protection for databases when the database is the product of
substantial investment but could not receive protection under copyright law.

Sui generis protection calls for a new right based upon economic grounds, it is not premised upon intellectual property principles that create a property right to offer protection to provide an incentive for further creation. This paper reviews the sui generis model for database protection to demonstrate its inherent failings that include its inability to foster technological or scientific innovation as well as its weakness from a Constitutional perspective. Databases are critical to all segments of society therefore it is important that the approach that one takes to regulate their development does not impair their vitality. Commodification of data and information is not the best way to structure competition for information goods. Taken together I believe that the protections that are discussed in Parts II, III and IV will provide the incentive to insure the continued development of global scientific, technological and academic communities. The Feist decision limited copyright protection to only original or creative “collections of information;” the bounds of this protection are examined in Part II. In Part III the misappropriation and contract theories of protection of information are reviewed and while the latter implicates the use of technology, it is in Part IV that the impact of technology on maintaining information’s exclusivity in fully explored. Technology which has burgeoned under the time honored intellectual property principles has created its own mechanisms for protecting databases; I suggest that left alone, the unregulated technology industry will continue to allocate private rights with the public interest to assure the progress of science necessary for the growth of society.

Part I: Sui Generis Doctrine

An Explanation
The sui generis protections of the E.U. Directive define a database as “collections of independent data…arranged in a systematic… and individually accessible” way (Directive 96/9 EC of the European Parliament and of the Council of March 11, 1996 of the Legal Protection of Databases O.J.). The European courts have interpreted this definition very broadly to include almost any grouping of data such as classified ads in newspapers, a collection of hyperlinks or the noncopyrightable data appended to an article (Maurer, Hugenholtz and Onsrud, 2001) so long as a substantial investment has been put into compiling the database. The E.U. Directive, and in essence of much of the proposed U.S. legislation, provide protection for the contents of a database which has been compiled with substantial investment from the “extraction” or “utilization” of “the whole or any substantial part” of the database if this would “conflict with the normal exploitation of the database,” undermine the “normal or potential market” or “unreasonably prejudice the legitimate rights of the rights holder.” The term of protection under the E.U. Directive is fifteen years from the time of creation and most of the proposed U.S. legislation followed a similar time line. Any substantial change that is made to the database is considered a new investment that starts the tolling of another fifteen years.

Negative Implications of Sui Generis
The practical constraints that result from the sui generis protection illustrate the defectiveness of this legal paradigm. A second comer would be unable to use information that is available on a website or published in a journal. Normally a second comer would freely download information from a website or request permission and perhaps pay a negotiated fee to use the data in the print example. The data under the sui generis paradigm is no longer in the public domain therefore a potential licensee can be refused. This will force the second comer to spend additional dollars, maybe obtain new financing, to repeat the data collection process possibly to obtain the identical data. This clearly contradicts common sense as well as the goals of knowledge production, which favor building upon previous discoveries and sharing information.
Many companies particularly in Europe are addressing the sui generis issue when establishing their own research policies. Fearful about committing millions of dollars to a research project that could be found to violate a sui generis right, companies advise employees not to use any data acquired by surfing the Net. Scientific progress depends upon combining data from various sources, under these legal constraints new discoveries will certainly be found more slowly.

Often times data is based upon one-time events that can not be physically regenerated; this is referred to as "sole source" data. Licensing of "sole source" data is often the optimal competitive solution to balance private rights and the public good. Under a sui generis scheme the compilers of "sole source data" are granted the same exclusive rights as all database owners who have invested substantial effort regardless of the uniqueness of the data elements. This forever excludes the world from the benefits of important information. In fact the original proposal of the E.U. Directive limited "unfair extraction" to commercial use which would have allowed for extraction for scientific and academic research thus leaving more information in the public domain. The limiting terms were disregarded and an optional exemption from protection for “teaching and scientific research” was put its place which could be adopted by the Member states.

Public sector data is not covered by sui generis protections but the source of funds for research is not always clear, thus diligent monitoring of funds dispersed by the government is needed to assure that the work product remains in the public domain. Many non-copyrightable databases that are generated by government funding are compiled by non-government employees at universities or in industry. In Europe, database protection has become a revenue generating opportunity for the government which efficiently takes information away from the public and puts into government coffers. Sui generis protection as legislated does not necessitate that the public is deprived of information but this is the likely result without government vigilance when funding projects.

The costs to society to acquire the data, which is granted monopoly protection under a sui generis regime, are manifold. There are the actual dollars spent to replicate previous research and the opportunity costs resulting from scientists' and other scholars' reluctance to use existing information. As a new legal regime develops that encourages commercial exploitation of research data sets, the cooperative environment that encourages interdisciplinary inspiration and innovation will be quashed. Therefore, placing a strong property right on databases which does not meet the threshold of creativity or originality will cause a major slow down in the high tech and scientific sectors which fuel the economy.

Constitutional Challenge to Sui Generis

Intellectual property rights in the United States are based in the Constitution in Article I Section 8 where Congress is granted the power to make laws, "To promote the Progress of Science and useful Arts, by securing for limited times…the exclusive Right to their respective…Discoveries." This clause of the Constitution has established a self-perpetuating paradox upon which intellectual property law has been forced to evolve. Normally, society relies upon competition to encourage growth and creativity, copyright, however, promotes monopolies to foster invention. The monopoly protection is stingily granted only when the threshold requirements of originality and creativity have been met. Professor Melville Nimmer (1991) an authority on both copyright and the First Amendment has explained that if it were not for the judicial exclusion of ideas and facts from the property rights granted to copyright authors and artists, the copyright protection would violate the First Amendment. The tension of the First Amendment and copyright is also relieved by the "fair use" doctrine which gives the second comer the opportunity to use the originator's creation for limited purposes which do not harm the originator's use. The purpose of copyright protection is not to create a monopoly to reward the originator with profits, therefore the "fair use" exemption or defense provides a balance between the second comer's First Amendment rights and the originator's copyright claims.
The sui generis protection is limited in the E.U. Directive to those activities which would conflict with the "normal exploitation" of the database and under the U.S. proposals the protection has been limited to harm that may be caused to "actual or potential markets." Limiting the reach of the database owners' control to prohibit the use of the second comer only if he interfered with the "actual or potential market" of the database owner was created to mirror the "fair use" exemption. Arguably, this was in order to legitimize or bestow a Constitutional blessing on the sui generis right. Noncreative databases do not rise to the level of original works of authorship and therefore do not merit Constitutional protection. The use of a copyright defense does not cure Constitutional flaws.

**Part II: Copyright Protection for Databases**

**Elements of a Database**

Databases comprise a variety of elements that have the potential to claim intellectual property protection, they are; the data elements, the research effort, the search or organizational tools used to access the data and the selection or arrangement of the data. In this Part II I show that how the time honored copyright principles that are based on encouraging creative productions of work can provide database protection.

The legal status of data elements is based upon the source and constituent parts of the data and although data may be separately categorized all data is considered facts for purposes of copyright law and facts belong in the public domain and do not receive copyright protection. Some data such as telephone numbers or statistics about sports or concert times are referred to as "synthetic data" because they are not derived from the outside world. Alternatively, there is data that is universally available through independent research. The database compiler may create the database himself or rely on another and seek the permission of the originator from whom he is sharing the information. Whatever rights are appurtenant to the data elements are passed through from the compiler to the end user.

The Supreme Court's 1991 decision in Feist Publishing, Inc. v Rural Telephone Service Co. , 499 U.S. 340(1991) made it very clear that the effort involved in finding and assembling a body of collected data was not entitled to copyright protection. There is debate in legal circles as to whether Feist's explicit renunciation of the "sweat of the brow" doctrine was breaking new ground or whether the decision was consistent with 200 years of American case law but its impact was enormous. The Feist Court rejected copyright protection for a white pages telephone directory thereby requiring originality or creativity in the actual compilation of the facts or data. As the decade of the nineties advanced and technology and the Internet progressed, the ramifications of the elimination of "sweat of the brow" from intellectual property rights protection took hold. Arguably, the Feist decision was the catalyst for the E.U. Directive and the cycle of legislative action that has been transpiring ever since.

Computer programs or software is the means by which data is organized and can be retrieved for use. Search engines are the most common Internet search or organizational tool but the success of many private database companies often depends upon the sophistication of similar tools. Fierce competition has characterized this field and the case of Computer Associates International, Inc. v Altai, Inc., 982 F.2d 693 (2d Cir. 1992) set the standard for copyright review. Since 1992 there has been universal acceptance of the "abstraction-filtration-comparison" test for finding intellectual property rights in software. In fact since the recent Federal Circuit's decision in State Street Bank & Trust Co. v Signature Financial Group, Inc., 149 F.3d 1368 (Fed. Cir. 1998) courts have found that software is patentable so long as it can produce a "useful, concrete and tangible result."
Data Arrangements after Feist

Discerning whether a compilation meets the requisite level of creativity for copyright protection after Feist is elusive. While the Feist Court repeatedly stated that the amount of originality needed is minimal, and that most compilations will pass the test, all of the databases in subsequent cases were not protectable. One of the more challenging arguments that a potential infringer may make to disqualify a database from copyright protection is based upon the "merger doctrine." The "merger doctrine" is based upon the idea-expression dichotomy. It is elusive because the original concept or contribution upon which the compiler designs his database is based upon an idea and ideas, like facts, belong in the public domain. An original concept or contribution which is deemed to be truly original will be protected but an idea must remain unprotected, i.e. in the public domain. This bind was clearly addressed in the Second Circuit when reviewing a compendium of projections of used car valuations, (CCC Information Services, Inc. v Maclean Hunter Market Reports, Inc., 44 F.3d 61, 2d Cir. 1994). The judge admitted that by viewing a compilation in terms of its idea alone the entire body of copyright law would be eviscerated, therefore he offered a way to view ideas that would yield certain ideas protectable and others not copyrightable.

An idea which is "infused with the author's taste and opinion" will provide the grounds or foundation for a database, but "an idea that undertakes to advance the understanding of phenomenon or the solution of problems" can not receive protection because it may materially assist the understanding of future thinkers. It is clear from the outset that to define an idea in terms of copyright law is but a legal conclusion based upon public policy concerns for stimulating creative works. In the case of databases ideas which tend to be functional or problem solving receive less protection than soft or subjective selections. This results from copyright's concern to keep the truth and knowledge in the open. The truth must be kept free, as we know because facts are the building blocks for knowledge and advancement.

Although intellectually inspiring, the "merger doctrine" is not the method most often used to evaluate the creativity of a database for purposes of copyright protection. Another guide for measuring the requisite activity is to judge the "thoughtfulness" of the selections. In Key Publications Inc. v. Chinatown Today Publishing Enterprises, Inc, 945 F.2d 50, (2d Cir 1991) a listing of New York's Chinese restaurants were compiled for the Chinese business community (Conley 1999). Another test is to look at how the compiler selected the data; this test was used in Warren Publishing, Inc. v. Microdos Data Corp. 115 F.3d 1509, (11th Cir 1999) where the subject matter was a Factbook of information about cable systems that was arranged by principal community. The Court of Appeals denied protection on several grounds that were due to lack of selection by Warren. Although Warren had rearranged the names of the cable systems, all systems listed by the FCC were included and Warren's arrangement was based upon a survey that had been conducted by cable operators not by Warren. Thus, the Factbook lacked originality.

Protection for databases under copyright is admittedly "thin" but the facts of Warren suggest that with a little more creativity millions of dollars of investment could have been saved. This case strongly suggests that in fact we do not need sui generis to protect the right of database compilers, rather post- Warren it should be a trivial matter for database manufacturers to add the necessary "spark" of originality to their product (Oram 2001).

Part III: Misappropriation and Contract Law Offer Protection

Misappropriation Doctrine

The digital environment lends itself to regulation under state misappropriation claims and database rights can be negotiated under contract. Both of these areas of law are clear and understandable; it becomes difficult to apply the legal protections when there is an overlap between state doctrine and Federal law. Relevant Federal law, which in this case is the Copyright Act, (17 U.S.C. 101 et. seq.) that covers the
same substantive area as state law will preempt state law, thus making it ineffective. It is not always clear whether both laws cover the subject matter.

The misappropriation doctrine provides a remedy for the compiler of information whenever another person duplicates or imitates the work at the compiler’s expense. The doctrine was devised by the U.S. Supreme Court to protect the rights of the Associated Press (AP) during World War I. (International News Service v Associated Press, 248 U.S. 215, 1918). The defendant International News Service (INS) had no reporters on the front therefore it would copy AP’s bulletins as they were published in New York and relay the news to their West Coast subscribers at the same time or even before the AP reached their West Coast subscribers. To fashion a remedy the Supreme Court granted a "quasi property " right to the first to disseminate news that one has invested in gathering. The data elements were facts and there was little creativity but the timeliness of news bestowed a quality that received protection.

The misappropriation doctrine was applied recently in the precedent setting case of National Basketball Association v Motorola, 105 F.3d 8941,( 2d Cir . 1997) where the plaintiff NBA attempted to prevent Motorola and Sports Team Analysis and Tracking Systems (STATS) from sending scores of ongoing games to users of Motorola's STATS paging device.

The plaintiffs were unable to stop Motorola based upon copyright principles because the actual plays of the game are not copyrightable and Motorola did not divert the broadcast feeds which are copyright material, rather it had its own reporters watching the games.

The NBA Court set out five elements of “hot news” which must be met in order to succeed in a misappropriation claim. They are: i) a plaintiff generates or gathers information at a cost; ii) the information is time sensitive; iii) a defendant's use of the information constitutes free riding on the plaintiff's efforts; iv) the defendant is in direct competition with a product or service offered by the plaintiff; and (v) the ability of others to free ride on the efforts of the plaintiff would reduce plaintiff’s incentive to produce the product.

The NBA was able to clearly satisfy the first two elements of the test and made a substantial case for the fourth. Motorola was not found to be free riding on NBA’s efforts however and there was no evidence that NBA’s incentive to play basketball or to broadcast the event (iv) was threatened by Motorola’s actions.

**Misappropriation and Preemption**

The Court’s analysis suggests that databases which lack any creativity may receive protection under the misappropriation doctrine where the compiler has a lead time advantage and the information is time sensitive. The key to garnering misappropriation protection depends upon the definition of “time sensitivity” and the level of competition between the compiler and the second comer. Dynamic compilations have various time sensitivities. A stock quote that is an hour old is worthless but a database of weather conditions may need to be updated only every two hours to be current. The time sensitive nature of “hot news” is what characterizes misappropriation and distinguishes it from similar copying activity, which would fall within the general subject matter of copyright. The Federal copyright statute only protects databases that are creative. The databases at hand lack creativity and they are therefore not protected by Federal copyright law. A state law that affords copyright protection equivalent to the Federal Copyright Act would be preempted. The Copyright Act preemption provision states that its rights are not equivalent to the state law however because the misappropriation doctrine involves “extra elements.” The “hot news” component of the misappropriation doctrine is the “extra element.” Therefore databases which contain time sensitive information will not be preempted and fall within the protection of the misappropriation doctrine, in fact, the more clear the time sensitivity the more persuasive the claim. Static databases remain beyond the protection of the misappropriation doctrine.
Contracts Afford Protection

Contracts are used to protect a database compiler’s rights in the digital age, most often in conjunction with technology to enforce those rights. Contract or more specifically, a license, is used by data owners to extract a fee from second comers and may take the form of a “shrink wrap” agreement when the information is held in a tangible form like a CD-ROM or a “click wrap” license when the information is obtained on line. The issue of the legal enforceability of such contracts was the subject of tremendous debate among numerous groups of policy makers, legislators, academics and industry experts. Finally after many failed attempts to enact uniform legislation, the Uniform Computer Information Transactions Act (UCITA) has been approved by the National Conference of Commissioners on Uniform State Laws and has been introduced and enacted in several states. UCITA has made it clear that licenses for electronic databases are enforceable. Licensees must be given notice of the terms and conditions for the use of the information and consent, i.e. the click on the computer screen, must be made affirmatively. When the user/licensee manifests his assent the terms form an enforceable contract.

With a contract a database owner can legally protect his rights to the compiled information despite, it’s lack of creativity or copyright. The issue of whether contracts which prohibit one from engaging in certain kinds of copying and distribution of a compilation is preempted by the Copyright Act was examined by Judge Easterbrook in the ProCD v Zeidenberg case 86 F.3d 1447, (7th Cir. 1996). The plaintiff, ProCD, had made an uncopyrightable compilation of telephone numbers, which he offered at two different prices, a more expensive commercial version and a personal use version. The defendant bought the personal use version under the shrink-wrap license, which forbid copying and began selling copies of parts of the database. The court upheld the terms of the contract enjoining Zeidenberg by finding that the contract was enforceable and that copyright rights are fundamentally different, because "a copyright is a right against the world. Contracts, by contrast, generally affect only their parties...(and) do not create 'exclusive rights", ProCD v Zeidenberg case 86 F.3d 1447, (7th Cir. 1996).The purpose of maximizing information is served by the contract because without the license protection the database owner would have been forced to sell at a single higher rate; a rate that would have been prohibitively expensive for the personal use buyer.

Part IV: Technological Protection and Its Implications

The technology, which has made information indispensable to millions of people, can also eliminate access. Encryption, which can take numerous forms, provides a very effective way for one to preserve his rights in a compiled database. Technologies are available that can prohibit copying information that is available on the Internet as well as techniques which prohibit reproduction of information that appear in video (DVD format) or music (SMDI format). These technical measures or self -help mechanisms may restrict the number of copies or forbid copying altogether. There are negative social implications associated with this technology if it makes access to knowledge prohibitive.

Information is a true public good in the strict economic sense because it is “non-excludable” and “non-rival.” It is “non-excludable” in that a creator of information may get a copyright on a book, for example, but that would not give him complete control over the extent to which a social benefit can be derived from it. Both the “first sale” doctrine and “fair use” principle of copyright law allow the buyer of the book to share the book with another, make limited photocopies and quote from the book. Information is also “non-rival” because one person’s consumption of it does not lessen its availability for another. Encryption makes it possible for holders of information to maintain complete control over access to information to exclude consumers who are not willing to pay a price that the compiler believes covers the investment in production. This is unsound economically because if a good (information) is non-rival its marginal cost is zero and since the compiler is charging a price to recoup costs then the good (information) which is being sold above its marginal cost is being underutilized (Benkler 2000).This trend may lead to the environment...
of perfect enclosure of information. Any degree of enclosure conflicts directly with the goals of establishing a regime for databases that will encourage the advancement of science and technology.

**Part V: Critical Assessment**

Throughout this paper knowledge development, technological and economic growth and advancement of science have all been used interchangeably to describe the goal of human society, and more specifically, to ascertain the intellectual property regime that will promote the development of databases. Databases have been shown to contribute to the progress of society. Incremental innovation which is central to the development of modern science and technology requires sharing of information. We must maintain the presumption that information is a public good and the existing fences that traditional intellectual property law have erected are sufficient to provide the necessary incentives to keep proprietors innovating.

Sui generis legislation does not encourage growth rather it rewards the data manufacturer. A study of several large commercial database companies in France, the U.K. and Germany illustrate that the E.U. Directive has not lead to sustained growth. Traditional copyright law will protect a database that contains a creative spark and online databases that are time sensitive will remain protected under the misappropriation doctrine. Reliance upon the market and a continued healthy technology sector which includes encryption will afford data producers protection against infringement.

I do not believe that any legislation which would increase the legal right of database compilers would enhance the progress of science. I favor a laissez faire approach leaving technology to balance the interests of the database creators and the second comers. I believe that price mechanisms offer the best solution within any limits the government may establish regarding the strength of the encryption. In suggesting reliance upon technology one must address the enclosure issue, however, which could be as harmful to the development of knowledge as the adoption of sui generis or stronger legislation. The implications of the sui generis model threaten the growth of knowledge in the United States to a greater extent than encryption in two ways. Once enacted the sui generis approach is applied universally and all second comers need to be aware and fearful of infringing on the originator's rights. This causes a chilling effect and reluctance to research freely. Secondly, although encryption is very effective it is imposed pursuant to a contract that is entered into voluntarily and rights are not imposed upon the whole world as under the sui generis regime.

Concerns about technology and copying go back to the days of the printing press yet with the basic interpretations the of Constitution, technology has flourished creating many astounding answers to life’s mysteries. In addition to technology, there are several legal mechanisms that can be drawn upon to protect databases, thus I would urge Congress not to import the sui generis right.

**References**


CCC Information Services, Inc. v Maclean Hunter Market Reports, Inc 44 F.3d 61 (2d Cir. 1994).


Biography

Julia Alpert Gladstone is a professor of Legal Studies at Bryant College in Smithfield Rhode Island, USA. She researches the relationship between technology and the law particularly as it is revealed on the Internet. For the past several years she has served as a chairperson for a Cyberspace Subcommittee of the American Bar Association as well as an editor of the Annual Cyberspace Law Survey. Her own writings cover a broad range of emerging areas of concern in cyberspace and she has been invited to speak in numerous locales around the globe.