Buffalo Law Review

Volume 51 | Number 1 Article 6

1-1-2003

Scientific Databases Should Be Protected under a *Sui Generis* Regime

Amol Pachnanda
University at Buffalo School of Law (Student)

COMMENT

Scientific Databases Should Be Protected Under a *Sui Generis* Regime

AMOL PACHNANDA†

INTRODUCTION

The publication of the entire human genome sequence in February 2001 marked a milestone in scientific achievement.[1] However, this remarkable moment was not without controversy.[2] Celera,[3] the bio-tech company that published the genome sequence, demanded that anyone wishing to ac-

† J.D. Candidate, SUNY Buffalo Law School, 2003. The author would like to thank the members of the Buffalo Law Review for their assistance in editing the drafts of this article.

1. *See* International Human Genome Sequencing Consortium, *Initial Sequencing and Analysis of the Human Genome*, NATURE, Feb. 15, 2001, at 860; J. Craig Venter et al., *The Sequence of the Human Genome*, SCIENCE, Feb. 16, 2001, at 1304.

2. The Human Genome Project (HGP) began in the 1980s following the United States Department of Energy (DOE) decision to create an ordered set of DNA segments from known locations to develop new computational methods for analyzing the genetic map and DNA sequence data. *See* Darryl R.J. Macer, 5 BIOETHICS 183 (1991).

3. Celera is a company dedicated to the discovery of new therapeutics using high technology. Celera's website can be accessed at http://www.celera.com (last visited Jan. 31, 2003). Celera has now embarked on an ambitious project in protein analysis. *See generally*, Robert F. Service, *Can Celera Do It Again?*, SCIENCE, Mar. 24, 2000, at 2136; Ellen Licking, *Beyond the Genome: Biotech's Next Holy Grail*, BUSINESS WEEK ONLINE, Apr. 10, 2000, *at* http://www.businessweek.com/2000/00_15/b3676117.htm?scriptFramed (last visited Jan. 31, 2003).

cess its genomic database first sign a licensing agreement contracting not to re-distribute the sequence to other users.[4]

This unique agreement disturbed[5] many within the scientific community because science generally functions on the free and open exchange of published data[6] as a means to build on pre-existing data, stir debate, cross-check and confirm published data. According to Craig Venter, leader of Celera's sequencing project, the reason for the unusual licensing agreement was that the United States does not allow copyright protection for databases.[7]

A database is a comprehensive collection of information easily organized to permit search, retrieval and organization, and Celera's unique arrangement illustrates the lack of copyright protection for databases in the United States today. To understand Celera's decision on a legal level, one

---

4. The unique agreement between Celera and Science allows academic users to download up to one megabase of Celera's sequence without restrictions. However, larger downloads require an agreement whereby redistribution is prohibited. Similarly, commercial users must sign a Material Transfer Agreement (MTA) agreeing to not commercialize their finding or redistribute the sequence. *See* Eugene Russo, *Behind the Sequence: Landmark Human Genome Papers Represent More Than Major Scientific Achievements*, SCIENTIST, Mar. 5, 2001, at 10. *See also* Declan Butler, *US/UK Statement on Genome Data Prompts Debate on 'Free Access,'* NATURE, Mar. 23, 2000, at 324.

5. "This is the first time in history that a paper reports a scientific result, but tells readers that to *see* it, they must sign a contract," says Eric Lander, director of the Whitehead Center for Genome Research. Russo, *supra* note 4, at 10. "We have patents, we have copyrights . . . but other than that we've never had a system for preventing people from using basic knowledge." *Id.* The publication, *Science*, also received a lot of criticism from the scientific community for publishing the sequence on Venter's terms. Science editor Donald Kennedy said, "It would be better in the best of all possible worlds if they had put the data in GenBank [the public genomic database]. . .But they couldn't do that because United States copyright protection for databases like this is just not adequate." *Id.* at 10.

6. *See* Tom Paulson, *Mapping Human Genome Reaches the End of the Road*, SEATTLE POST-INTELLIGENCER, Feb. 12, 2001, *available at* http://seattlepi.nwsource.com/local/geno12.shtml (last visited Feb. 15, 2003).

7. *See* Arti Kaur Rai, *Regulating Scientific Research: Intellectual Property Rights and the Norms of Science*, 94 N.W. U. L. REV. 77 1999. According to some industry observers, however, what Venter is doing is "not a commercial venture. It's really Craig Venter going after the Nobel Prize for sequencing the genome." *Id.* at 115. In addition to the issue of copyright protection for databases, there are also questions regarding patents on human DNA sequences. *See, e.g.*, Donna M. Gitter, *International Conflicts Over Patenting Human DNA Sequences in the United States and European Union: An Argument for Compulsory Licensing and a Fair-Use Exemption*, 76 N.Y.U. L. REV. 1623 (2001).

has to go back to the line of cases that eradicated the protection of databases. Any discussion about the protection of databases must begin with *Feist Publications v. Rural Telephone Service Co.*[8] In *Feist*, the Supreme Court unanimously held that copyright protection only encompassed compilations displaying some degree of "creative originality" in their selection, coordination, and arrangement of facts, and that investment of money-"sweat of the brow"-did not, by itself, merit copyright protection.[9] The decision left database compilers, especially those compiling factual databases, with an extremely "thin" layer of protection.[10]

Generally, creative databases can meet the low threshold requirement of creativity, usually by incorporating the producer's subjective thoughts. In contrast, factual compilations are most beneficial when they conform to more uniform standards and allow end users to manipulate the data to their own advantage. As such, factual compilations cannot meet even this low threshold of creativity. Consequently, Celera was reluctant to publish the raw human genomic sequence,[11] a factual compilation, without restricting the ability of others to make money by redistributing the data without incurring the associated costs of production and organization.

While *Feist* and its progeny have threatened database protection in the United States, the European Union has decisively addressed this issue. In 1996, the European Union issued a Directive[12] which conferred *sui generis*[13]

8. 499 U.S. 340 (1991).

9. *See id.* at 359-60 ("Originality, not 'sweat of the brow,' is the touchstone of copyright protection").

10. *Id.* at 361. For an overview of the traditional forms of intellectual property used to provide legal protection for information technology, see Lionel M. Lavenue, *Database Rights and Technical Data Rights: The Expansion of Intellectual Property for the Protection of Databases*, 38 SANTA CLARA L. REV. 1 (1997).

11. The raw genomic sequence is only the first step towards the development of commercially viable products. Researchers still face the long task of analyzing the raw sequence to determine which parts actually encode for genes, determine the sequences that regulate transcription, and ultimately determine the products the genes encode. *See* Alexander K. Hass, *The Wellcome Trust's Disclosures of Gene Sequence Data into the Public Domain & the Potential for Proprietary Rights in the Human Genome*, 16 BERKELEY TECH. L.J. 145, 146-47 (2001).

12. The text of the Directive is available at http://eon.law.harvard.edu/property00/alternatives/ directive.html (last visited Feb. 14, 2003).

13. *Sui Generis* means "of its own kind, peculiar." BALLENTINE'S LAW

protection to databases for which a substantial investment of time, effort, or money was put in for their production. Following the European Union's lead, four separate pieces of legislation were proposed in the United States with regards to the creation of a *sui generis* protection right for databases.[14] The two latest proposals were the Collections of Information Antipiracy Act and the Consumer and Investor Access to Information Act of 1999. However, neither of these bills have received Congressional approval.

This Comment argues that the current state of copyright protection for databases in the United States is inadequate and must be addressed by Congress. Part I discusses the current state of protection for databases, and includes an analysis of the *Feist* decision and subsequent database cases. Part II examines the importance of databases in furthering the scientific endeavor by increasing the accessibility and accuracy of information. Specific attention will be paid to scientific databases and databases arising out of the Human Genome Project. Also, attention will be given to the lack of adequate protection for electronic databases.[15] Part III examines the response of the European Union to the decision in *Feist*. Part IV examines the United States' failed attempts to extend *sui generis* protection to databases. Finally, Part V examines the inadequacy of state law alternatives for protecting databases. Ultimately, it will become clear that the gap created by the Copyright Act and *Feist* must be filled by legislative action.

DICTIONARY 1236 (3d ed. 1969).

14. The four proposals, starting in 1996 are: Database Investment and Intellectual Property Antipiracy Act, H.R. 3531, 104th Cong. (1996); Collections of Information Antipiracy Act, H.R. 2652, 105th Cong. (1998); Collection of Information Antipiracy Act, H.R. 354, 106th Cong. (1999); and The Consumer and Investor Access To Information Act of 1999, H.R. 1858, 106th Cong. (1999).

15. Recent technological advancements in on-line piracy and increased availability of electronic databases allow users to duplicate and redistribute contents of a database with impunity.

I. THE CURRENT STATE OF DATABASE PROTECTION

A. *The Law Before* Feist Publications, Inc. v. Rural Telephone Service Co.

Prior to the Supreme Court's 1991 decision in *Feist*,[16] database compilers sought protection for their databases under the "sweat of the brow" doctrine.[17] Under this theory compilations resulting from a substantial effort were granted full copyright protection. As a result, future compilers were prohibited from copying facts from the protected compilations, effectively eliminating the competitive advantage of second-comers. However, the monumental decision in *Feist* eliminated the protection afforded to compilers for their investment and for the time spent collecting information and placing it in a useable form.[18]

B. Feist Publications, Inc. v. Rural Telephone Service Co.

Rural Telephone Service Company (Rural) published a typical telephone directory listing the names of its subscribers in alphabetical order.[19] Feist Publications, Inc. (Feist Publications), published an area-wide directory that cov-

16. 499 U.S. 340 (1991).

17. *See* Ill. Bell Tel. Co. v. Haines & Co., 905 F.2d 1081 (7th Cir. 1990); Hutchinson Tel. Co. v. Fronteer Directory Co. of Minn., 770 F.2d 128 (8th Cir. 1985). However, the majority of circuits applied the "creative selection" doctrine, which only conferred protection to the creative aspects of databases. *See* Worth v. Selchow & Righter Co., 827 F.2d 569 (9th Cir. 1987); Southwestern Bell Tel. & Tel. Co. v. Associated Tel. Directory Publishers, 756 F.2d 801 (11th Cir. 1985); Eckes v. Card Price Update, 736 F.2d 859 (2d Cir. 1984); Miller v. Universal City Studios, Inc., 650 F.2d 1365 (5th Cir. 1981); *See also* Alfred C. Yen, *The Legacy of Feist: Consequences of the Weak Connection Between Copyright and the Economics of Public Goods*, 52 OHIO ST. L.J. 1343, 1344 (1991) (discussing the "creative selection" doctrine); Tracy Lea Meade, Note, *Ex-Post Feist: Applications of a Landmark Copyright Decision*, 2 J. INTELL. PROP. L. 245, 250 (1994). *See generally* Jordan M. Blanke, *Vincent Van Gogh, "Sweat of the Brow," and Database Protection*, 39 AM. BUS L.J. 645 (2002).

18. *See* Paul Goldstein, 38 J. COPY. SOC'Y 109, 118 (1991). Despite the impact of the *Feist* decision, its importance has not been free from question. "I think everyone is giving way too much importance to *Feist*. *Feist* was a unanimous Supreme Court opinion, which usually means a lot in constitutional law. In intellectual property law, it means half of the justices were asleep and didn't care about the case." Symposium, *Database Protection*, 11 FORDHAM I.P. MEDIA & ENT. L.J. 275, 294 (2001) (Comments of Professor Hansen).

19. 499 U.S. at 342.

ered a much larger geographic area. The larger directory eliminated the need to consult multiple directories.[20] Feist Publications agreed to pay each of the eleven local telephone companies for the right to use their listings.[21] Of the eleven, only Rural declined.[22] Regardless, Feist Publications used Rural's white-page listings without its consent.[23]

Feist Publications eliminated listings that fell outside the scope of its directory, hired employees to investigate the listings that remained, and further verified the data recorded by Rural.[24] Despite modifications, 1,309 of the 46,878 listings in Feist Publications' directory were identical to the listings in Rural's white pages.[25]

Rural sued for copyright infringement, taking the position that Feist Publications could not use the information in Rural's white pages to compile its own directory, but must independently collect the data.[26] Feist Publications' response was that conducting its own research would be economically unfeasible, wasteful and also unnecessary since the information in Rural's white pages was not protected by copyright.[27]

Writing for the Supreme Court, Justice O'Connor addressed the undeniable tension posed by copyright's protection for compilations,[28] but not facts.[29] According to O'Connor, originality, the "*sine qua non* of copyright,"[30] requires independent creation by the author plus some minimal degree of creativity.[31] The requisite level of creativity is extremely low and most works "no matter how crude, humble or obvious," will qualify.[32]

---

20. *Id.* at 343.
21. *Id.*
22. *Id.*
23. *Id.*
24. *Id.* at 343-44.
25. *Id.* Included were false listings Rural had inserted to detect copying. *Id.*
26. *Id.*
27. *Id.* The district court granted summary judgment in favor of Rural, holding that the directories were copyrightable. The Tenth Circuit affirmed. *Id.*
28. Many compilations contain nothing more than raw data wholly devoid of any written expression.
29. The issue before the Supreme Court was whether Rural's copyright in the directory extended to the names, telephone numbers, and towns copied by *Feist*. *See Feist*, 499 U.S. at 344.
30. *See id.* at 345.
31. *Id.* at 346.
32. *Id.* at 345.

O'Connor further pointed out that originality is a constitutional requirement.[33] The originality requirement mandates the disparate treatment of facts and compilations. Facts are not created and as such, do not owe their origin to an intellectual conception of an author.[34] The discoverer of facts simply finds the information and copies it. The distinction, according to O'Connor, is that between creation and discovery.[35]

Justice O'Connor emphasized that the copyright protection for compilations is extremely thin.[36] Some subjective input of the author is required to merit copyright protection for the compilation. However, subsequent users are free to copy the underlying facts if they select, arrange, or coordinate the facts in a "creative" way. According to O'Connor it is not unfair that second-comers may freely use the fruits of the first compiler's labor without compensation.[37] This is the essence of copyright—to promote the progress of science and the arts by allowing others to freely build upon ideas and information contained in pre-existing works.[38]

Applying the doctrine to the facts, O'Connor explained that Feist Publications, by copying the names, towns, and telephone numbers of Rural's subscribers, took nothing that was "original" to Rural.[39] The raw subscribers' information were facts discovered by Rural and did not originate from an intellectual creation.[40] According to O'Connor, "'[t]he originality requirement' rules out protecting . . .names, addresses, and telephone numbers of which plaintiff by no stretch of the imagination could be called author."[41] Rural's

33. *Id.* at 346. The opinion points to two decisions that articulated the idea that originality is the touchstone of copyright protection. *See* The Trade-Mark Cases, 100 U.S. 82 (1879) (the court explained that originality requires independent creation plus minimal degree of creativity); Burrow-Giles Lithographic Co. v. Sarnoy, 111 U.S. 53 (1884) (copyright is limited to the original intellectual creations of the author).

34. *Feist*, 499 U.S. at 347.

35. *See id.*

36. *See id.* at 349.

37. *Id.*

38. This theory is known as the idea/expression dichotomy and applies to all works of authorship.

39. *Feist*, 499 U.S. at 361.

40. *Id.*

41. *Id.* (quoting Patteron & Joyce, *Monopolizing the Law: The Scope of Copyright Protection for Law Reports and Statutory Compilations*, 36 UCLA L.

white pages were "typical," "garden-variety" and entirely devoid of even the minimal hint of creativity[42] necessary to meet copyright's originality requirement.[43]

The *Feist* decision established that the protection for databases, constitutionally and statutorily, was weak. According to O'Connor, courts that permitted protection of database under the "sweat of the brow" theory had misinterpreted the Constitution.[44]

C. *The Impact of* Feist Publications, Inc. v. Rural Telephone Service Co.

Feist made it clear that only a compiler's selection and arrangement of facts would be protected, while the raw data could be copied at will. With such a "thin"[45] layer of protection, the threat of piracy may discourage the development of commercially valuable databases. Database producers will be reluctant to spend the time and money necessary to compile a useful database if competitors can copy and exploit it easily for their own profit.[46]

Since *Feist*, courts have analyzed database cases on two levels.[47] First, the court determines if the author's subjec-

---

REV. 719, 776 (1989).

42. *Id.* at 362.

43. *Id.*

44. *Id.* at 353-54.

45. *Id.* at 349.

46. W. Matthew Wayman, Comment, *International Database Protection: A Multilateral Treaty Solution to the United States' Database Dilemma*, 37 SANTA CLARA L. REV. 427, 435 (1997).

47. "The *Feist* decision began the slide down a slippery slope of deteriorating protection for database providers." Russell G. Nelson, *Recent Development: Seeking Refuge From a Technology Storm: The Current Status of Database Protection Legislation After the Sinking of the Collections of Information Anti-Piracy Act and the Second Circuit Affirmation of Matthew Bender & Co v. West Publishing Co*, 6 J. INTELL. PROP. L. 453, 461 (1999). Subsequent court decisions have created an analytic structure that can be summarized as follows:

> (1) The mere amassing of data, even if with innovative technique, is not copyrightable; (2) What makes a compilation of data copyrightable is the selection, coordination, or arrangement of those data. Without at least one of these elements no database will be copyrightable. (3) The selection or arrangement needed to secure a copyright must not only be original but also "creative" (4) Selection and arrangement must have an element of subjectivity they must embody the judgment of the compiler. (5) Selection and arrangement can occur at either the macro level or the micro level; (6) Selection or arrangement will not be protected to the extent that the resulting database has functional utility (7) selec-

tive views were incorporated into the selection and arrangement of the facts. Second, it appears that, regardless of whether this was the case, courts have held there can still be no copyright infringement. Two circuit court cases elucidate how truly thin copyright protection is for factual databases.

In *Key Publications, Inc. v. Chinatown Today Publishing Enterprises, Inc.*,[48] the plaintiff sued the defendants for infringing its copyright in the Chinese-American yellow pages directory.[49] The district court held that the defendant had infringed the plaintiff's copyright in the directory.[50] The Second Circuit held that the directory possessed the requisite originality because the plaintiff had exercised independent judgment in deciding which information to include, and the order in which to include it.[51] However, since the defendant's selection and arrangement of the information contained in its directory was different than the selection and arrangement of the plaintiff's directory, there was no copyright infringement.[52] Even though the court held

---

> tion and arrangement at the macro level must be closely scrutinized for merger of idea and expression [and] (8) infringement of a database will be judged by comparing the selection or arrangement of the two works, not the data themselves. Copying of data is not an infringement if the selection and arrangement in defendant's work are not substantially similar to the selection and arrangement in plaintiff's work.

Id.

48. 945 F.2d 509 (2d Cir. 1991).

49. *Id.* at 511.

50. *Id.* at 511-12

51. *Id.* at 513.

52. *Id.* at 515-17. Whether a database receives copyright protection is ultimately dependent of the nature of the facts in the database. In *Feist*, the facts (names, telephone number, and addresses of residents in a particular area) existed in nature. These facts were not comprised of anything original to the author. By way of contrast, the facts in *CCC Information Services, Inc. v. MacLean Hunter Market Reports*, were original to the author because the valuation of cars was derived from a unique scheme. 44 F.3d 61 (2d Cir. 1994). In *CCC Information Services*, the author, Maclean Hunter, published a book setting forth the author's projections of the valuations of used cars. *Id.* at 63. The valuations represented the author's professional judgment and predictions based on other information sources. *Id.* The Circuit Court firmly rejected the lower court's conclusion that the used car valuations, like the telephone numbers in *Feist*, were pre-existing facts. *Id.* at 67. The valuations were not reports of historical prices, numbers which would be pre-existing facts. The valuations were original creations of Maclean because the number was derived by weighing several factors. *Id.* Facts in genomic databases are more closely aligned to the concept of facts in *Feist*, and therefore will not merit the greater protection afforded in *CCC*

that the database compiler's arrangement and selection was based on subjective choices, only the arrangement and selection was protected, not the underlying facts.[53]

Perhaps the most disturbing post-*Feist* case is *Warren Publishing v. Microdos Data Corporation*,[54] where the Court held that cable listings were not original enough to deserve protection. The *Warren* decision is an example of inequitable decision making. It rewards copying, not creativity. Warren Publishing was the publisher of a directory which provided information on cable television systems in the United States.[55] The defendant, Microdos Data Corporation (Microdos), also marketed a computer software package containing a compilation of facts about cable systems.[56] Warren Publishing brought a copyright infringement suit alleging the Microdos's computer software compilation infringed its cable access directory by copying its selection and arrangement.[57] On these facts, the district court enjoined Microdos from violating Warren Publishing's copyright, holding that the community system utilized by Warren Publishing was based on its unique definition of a cable system and was "sufficient[ly] creative and original to be

---

Information Services.

In several other cases, the courts held that selections of facts were copyrightable. *See* Nester's Map & Guide Corp. v. Hagstrom Map Co., 796 F. Supp. 729 (E.D.N.Y. 1992) (holding that the selection of both street address listings and out of town destinations in the New York City's taxi drivers guide were copyrightable); Kregos v. Associated Press, 937 F.2d 700 (2d Cir. 1991) (court held that selection of nine categories for use in a baseball pitching form used to predict the outcome of games demonstrated sufficient originality to merit protection). Cases in which selection was held not to be creative include Bellsouth Adver. & Publ'g Corp. v. Donnelly Info., 999 F.2d 1436 (11th Cir. 1993) (holding that the selection of a geographic scope and closing date for a yellow pages directory was not original); Victor Lalli Enterprises v. Big Red Apple, 936 F.2d 671 (2d Cir. 1991) (holding that a chart containing thirteen months of winning numbers in New York City's illegal numbers game was not entitled to copyright protection); .

53. *See Key Publ'n, Inc.*, 945 F.2d at 512. This is in contrast to *Feist* where the court held that the selection and arrangement was completely devoid of originality.

54. 115 F.3d 1509 (11th Cir. 1997). For a discussion of that decision, see Andrew Oram, *The Sap and the Syrup of the Information Age: Coping with Database Protection Laws*, *available at* http://www.praxagora.com/andyo/ professional/collection_law.html#footnote_42 (last visited Feb. 15, 2003).

55. *Warren Publ'g*, 115 F.2d at 1511.

56. *Id.* at 1512.

57. *Id.*

copyrightable."[58] The selection of communities utilized by Microdos was found to be "substantially similar" to that of Warren Publishing.[59]

The Eleventh Circuit disagreed, finding that Warren's method of gathering information for its compilation lacked originality and was excluded from copyright protection. The Court held that Warren Publishing failed to show that it had exercised any judgment in "selecting" which cable systems to include, "but rather included the entire relevant universe known to it."[60] The court further explained that, even if it were to assume that Warren Publishing's selection of principal communities was creative and original, it would still not be entitled to copyright protection "because the selection [was] not its own but rather that of cable operators."[61] Warren Publishing's techniques of contacting the cable operators to determine which communities were considered lead communities were "not acts of authorship, but techniques for the discovery of facts."[62] The Court noted that, "[j]ust as the Copyright Act does not protect 'industrious collection,' it affords no shelter to the resourceful, efficient, or creative collector."[63] On this basis the Court held that, though Warren may have found an efficient method of gathering information, it lacked sufficient creativity and originality to be entitled to copyright protection.[64]

Thus, in the cases since *Feist*, courts have consistently held that there is a "thin" layer of protection for databases. Even if creativity and originality is found in the selection and arrangement of the data, there is nothing to prevent wholesale copying of the underlying facts. These decisions clearly demonstrate that copyright protection for databases is essentially non-existent in the United States.[65]

---

58. *Id.* at 1513.

59. *Id.*

60. *Id.* at 1518. According to the Court, the only decision it did make was used to make the directory more commercially useful. *Id.*

61. *Id.* at 1519.

62. *Id.*

63. *Id.* at 1520.

64. *Id.*

65. *See* Paula Baron, *Back to the Future: Learning from the Past in the Database Debate*, 62 OHIO ST. L.J. 879, 887 (2001) (proposing that a *sui generis* scheme based upon the principles of old copyright cases, coupled with a shorter period of protection than that is available under copyright law, could provide the proper balance between access to information and incentive to produce more information).

II. THE IMPORTANCE OF DATABASES IN FURTHERING SCIENCE

A database[66] is a comprehensive collection of related information easily organized to permit search, retrieval and reorganization. Under the 1976 Copyright Act,[67] databases are protected as "compilations." The Act defines a compilation as "a work formed by the collection and assembling of preexisting material or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship."[68] A critical element of this definition is the idea that the underlying material or raw data is not protected.[69] It is the selection, coordination, or arrangement of that data that is the subject of copyright protection. As such, copyright law does not prevent a third party from extracting and re-utilizing the data from a compilation, even if the particular selection and arrangement merits protection.

However, the most useful databases are those that contain comprehensive, organized, and current information.[70] To receive even the minimal protection for selection and arrangement of the database, a database producer must create a database that is arranged without any logic, and therefore hard to use in order to receive protection from infringement.[71] This lack of protection is incongruous with

---

66. Databases are the tools that provide information about information and have become the new building blocks of knowledge. Michael J. Bastian, Note, *Protection of "Non-creative" Databases: Harmonization of United States, Foreign, and International Law*, 22 B.C. INT'L & COMP. L. REV. 425, 426 n.9 (1999).

67. 17 U.S.C. § 101 (2000).

68. *Id.* For definitions of terms that consistently appear in discussions of compilations or databases, see Stacey H. King, *Are We Ready to Answer the Question?*: Baker v. Selden, *The Post Feist Era and Database Protection*, 41 J.L. & TECH 65 (2001).

69. "A striking example of the harm that can result from a grant of proprietary rights in data is the privatization of data from the Landsat series of remote sensing satellites. . . .Following privatization, the prices of Landsat data increased from approximately $400 to $4,400 per image." Bastian, *supra* note 66, at 431.

70. *See* John F. Hayden, *Recent Development: Copyright Protection of Computer Databases After Feist*, 5 HARV. J.L. & TECH. 215, 229-30 (1991).

71. Lisa Barr, *Legislative Update: Database Protection Bill*, 8 J. ART. & ENT. L. 371, 376-77 (1998). Database producers have responded to *Feist* by adding copyrightable elements to their databases. Consequently, these databases are more expensive to produce and the information contained in them is not readily accessible. *See* Jason R. Boyarski, Note, *The Heist of Feist: Protection for Collections of Information and the Possible Federalization of "Hot News,"* 21 CARDOZO L. REV. 871, 905-06 (1999). Under *Feist*, copyright protection is not contingent

the demonstrated demand and utility of databases.[72] Databases are a vital part of virtually every segment of the economy.[73] Though the value of databases to society is high, so is the cost of producing comprehensive databases.[74] Their widespread use and critical importance necessitates a form of legal protection against their misappropriation.[75]

The internet has played a large role in increasing the accessibility of databases. At the same time, however, it has increased cost and protection concerns of database compilers.[76] Database users can quickly migrate from one database to another and can quickly acquire information. Once a database is put on-line, the concerns over unauthorized access, duplication of content, and mass redistribution of information produced with large investments of time, money, and effort, are greatly exacerbated.[77] The concern is that second-comers can have access to, and profit from, purely factual information at no cost.

Scientific databases, usually available online, epitomize this problem. Few databases will meet the low threshold of creativity because their utility arises from the fact that they are exhaustive and standardized.[78]

---

upon the quality, or usefulness of the compilations, but rather entirely upon the creativeness of the presentation to the user. *See* Paul Durdik, *Ancient Debate, New Technology: The European Community Moves to Protect Computer Databases*, 12 B. U. INT'L L.J. 153, 176 (1994).

72. For a discussion regarding the general importance of databases see Jeffrey C. Wolken, Note, *Just the Facts, Ma'am. A Case For Uniform Federal Regulation of Information Databases in the New Information Age*, 48 SYRACUSE. L. REV. 1263, 1266 (1998).

73. *See* Hayden, *supra* note 70, at 215-16.

74. The primary cost of a database is the high production cost. The production cost consists of both data preparation and data collection. The preparation cost is associated with preparing a database, ensuring data quality, accuracy, and enhancing the utility of the database for the end user. *See* A QUESTION OF BALANCE: PRIVATE RIGHTS AND THE PUBLIC INTEREST IN SCIENTIFIC AND TECHNICAL DATABASES 44 (1999) [hereinafter A QUESTION OF BALANCE].

75. *See id.*

76. Digital technology had greatly enhanced a database producer's ability to compile and disseminate data. It has also enhanced the ability of second comers to cheaply copy or manipulate the contents and disseminate the information to other users at a much lower cost. *See* Bastian, *supra* note 66, at 429.

77. "Cost recoupment in itself, however, is not a goal of copyright policy unless protecting the compilations at issue also promotes the progress [of science] in some way." Denise R. Polivy, *Feist Applied: Imagination Protects, but Perspiration Persists: The Bases for Copyright Protection for Factual Compilations*, 8 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 773, 800 (1998).

78. *See* Durdik, *supra* note 71, at 169.

A. *Scientific Databases and Factual Compilations in General*

Factual data is the essential building block of scientific research.[79] As the complexity of scientific research continues to increase exponentially, so does the need for larger and more comprehensive sets of databases. In the course of research, scientists frequently draw on existing databases in order to use information to provide new insights and advance our understanding of the natural world. This is an essential part of the scientific process.[80]

Scientific databases present three unique problems and needs: the need for monetary incentives for commercial companies to invest in commercializing databases; the need to provide access to academic users at a low cost to further scientific research; and the scientific community's need for full and unrestricted access to data.[81]

The genomic databases[82] have been the voice behind a new form of protection because of their unique nature. The genome sequence data differs from other forms of biological data, because it is the fundamental data that will be used by scientists for centuries. Therefore, a mode of protection that will ensure its continued availability and accuracy is imperative.[83]

---

79. *See* Polivy, *supra* note 77, at 777.

80. International Council of Scientific Unions, Committee on Date for Science and Technology Group on Data and Information, *Position Paper on Access to Databases*, *available at* http://www.codata.org/data_access/wipo.pdf (last visited Feb. 15, 2003).

81. The later problem inevitably conflicts with "self-help" strategies based on secrecy or partial disclosure. *See* Stephen M. Maurer, *Raw Knowledge: Protecting Technical Databases for Science and Industry* 1, *available at* http://www.sims.berkeley.edu/courses/is296a-3/s99/database.pdf (last visited Feb. 15, 2003).

82. *See generally*, Michael Brudno, *Protection of Genomic Databases*, *available at* http://www.stanford.edu/~brudno/essays/genbase.pdf (last visited Feb. 15, 2003).

83. For opinions contending that there should be no right in the human genome because it is the Common Heritage of Humanity, see Paige Gardner, *Technology and the Environment: Mapping of the Human Genome: Current Approaches to Sharing the Benefits With Developing Countries*, 2000 COLO. J. INT'L ENVTL. L. & POL'Y 281 (2000); J.M. Spectra, *The Fruit of the Human Genome Tree: Cautionary Tales About Technology, Investment, and the Heritage of Humankind*, 23 LOY. L. A. INT'L & COMP. L. REV. 1 (2001); Melissa L. Sturges, Note & Comment, *Who Should Hold Property Rights to the Human Genome? An Application of the Common Heritage of Humankind*, 13 AM. U. INT'L L. REV. 219 (1997).

The next twenty years will bring tremendous technological innovations, such as cures for diseases, bioinformatics, and improved computer technology. Maybe the most important innovation in science will lie in the area of content/information.[84] A legislative regime to protect the investment in information and ensure continued flow of information is critical.[85]

B. *Databases in the Human Genome Project*

In May 1995, an unknown virulent strain of the microbe that causes tuberculosis infected a twenty-one year old factory worker. The final tally of infected individuals totaled two-hundred and twenty people, including 75% of his co-workers and 80% of his acquaintances.[86]

Robert Fleishmann of the Institute for Genomic Research (TIGR), a non-profit genetics research group in Maryland agreed to sequence the genome of the microbe and make it available on-line six months later.[87] The delay was not to verify results, but instead to give TIGR's commercial partner, Human Genome Sciences (HGS), an early look at new discoveries and the time to seek intellectual property rights.[88]

This agreement rekindled the debate about who should control DNA sequence data and how quickly it should be shared. One group of researchers argued that failure to share data quickly leads to duplication and waste.[89] By way of contrast, researchers at commercial enterprises opposed the idea of immediate data release.[90] They argued that immediate release is analogous to asking a pharmaceutical company to give away its formula for a drug.[91]

---

84. *See* Bastian, *supra* note 66, at 431.

85. Absent reform from Congress, traditional forms of intellectual property protection are inadequate to protect the investments made by genomics companies in acquiring genome data. However, protection might be unnecessary because some companies can still benefit from information in the public domain by providing access that is easier and more efficient. *See* Hass, *supra* note 11, at 162.

86. Eliot Marshall, *Ethics in Science: Is Data-Hoarding Slowing the Assault on Pathogens?,* SCIENCE, Feb. 7, 1997, at 777.

87. *Id.*

88. *Id.*

89. *Id.*

90. *Id.*

91. *Id.* In this instance, TIGR was a facility funded by both public and pri-

The human genomic sequence is a "natural database, transforming biological research and applications over the past decade into a data dependent enterprise."[92] The raw data needed to conduct basic genetic and biotech research is maintained in nearly two-hundred public sector databases throughout the world.[93] From a computing perspective, many of these sites tend to be outdated, inaccurate, and non-standardized.[94]

The best known of these public databases is GenBank.[95] However, even GenBank's capabilities are limited. For example, the available search tools can only perform full word searches, while editing and commenting is strictly confined to the author's annotations. No effort is made to comment on related journal articles or resolves conflicts. Moreover, it is virtually impossible to update sequences.[96] To exacerbate the problem, several non-profit biotech databases have been forced to shut-down.[97] Irate researchers have conceded that the community may have to "get by" with inadequate updating, editing, and annotations.[98]

GenBank is a perfect example of the sometimes low quality of information in public databases.[99] Private databases tend to be more comprehensive because, generally, they have been repeatedly checked. In a profession where experiments last months, a crucial mistake can set a team of researchers back in their quest.[100]

---

vate dollars. *Id.* However, if public dollars were used to fund the research, then that information should have been made available as soon as possible. If TIGR was the only facility that could have decoded the genome then it exercised a monopoly over public data. This is inapposite to the way science functions.

92. A QUESTION OF BALANCE, *supra* note 74, at 17-18.

93. Maurer, *supra* note 81, at 14.

94. *Id.* at 15.

95. *Id.* GenBank provides researchers with tools to hunt for new genes, compare the evolution of genes in different organisms, and determine functions of new genes. *See* Elizabeth Pennisi, *Keeping Genome Databases Clean and Up to Date*, SCIENCE, Oct. 15, 1999, at 447.

96. Maurer, *supra* note 81, at 15.

97. *Id.*

98. *Id.*

99. David Lipman, director of the National Center for Biotechnology Information, admits that NCBI's Genbank has its limitations. "It does not represent what we know of biology at any given time [because] [i]t only represents what the author put in." Declan Butler & Paul Smaglik, *Draft Data Leave Geneticists With a Mountain Still to Climb*, NATURE, Jun. 29, 2000, at 984, 985.

100. *See* Elizabeth Pennisi, *Keeping Genome Databases Clean and Up to Date*, SCIENCE, Oct. 15, 1999, at 447.

The primary advantage of private databases is that they are usually well maintained and better annotated.[101] Private database producers require more than the clerical entry. For example, researchers are required to submit relevant journal articles, and cross-reference genetic information from other organisms that are related to the entry of sequences.[102] Private researchers must also sort through conflicting data because, at times, researchers use different terminology to describe the same sequence.[103] By way of contrast, GenBank often contains entries that appear different but actually refer to the same sequence.[104]

The rate of scientific progress depends in equal part on the continued collection of new data, the accuracy of the data collected, and the dissemination of the information in the new databases.[105] This requires both time and money. The database must be organized with efficient structure, presentation, and format; provide complementary analytical support software; and provide optimal quality assurance.[106] As a database becomes larger and more complex, database maintenance becomes increasingly important and constitutes a significant component of the overall cost of the database. This applies to the human genome project. The databases that will result from the genomic sequences of different organisms will extend far beyond just the raw sequence data.[107]

Since the DNA sequence of each organism exists in nature, the sequencer is merely a discoverer. Without the "sweat of the brow" doctrine, there is no copyright protection for DNA sequences.[108] If a researcher constructs

---

101. Brudno, *supra* note 82, at 8.

102. *Id.*

103. *Id.*

104. *Id.*

105. *See* A QUESTION OF BALANCE, *supra* note 74, at 26.

106. *Id.* at 27.

107. Genome related databases are broken down into generalized and specialized databases. "Generalized databases include the GenBank/EMBL/DDBJ archives of nucleic acids sequences and the PIR and SwissProt polypeptide sequence databases." William M. Gelbart, *Databases in Genomic Research*, SCIENCE, Oct. 23, 1998, at 659. These databases are important in presenting information about molecules without their functional characteristics. By way of contrast, specialized databases are concentrated around a specific organism or specific biological function. *Id.*

108. James G. Silva, *Copyright Protection of Biotechnology Works: Into the Dustbin Of History?*, 2000 B.C. INTELL. PROP. & TECH. F. 012801 (2000) *avail-*

a sequence based on a sequence that exists in nature, but was undiscovered, then he has not created anything original.[109] If a researcher then arranges the sequence in a creative way, the database has lost its utility. While it has been postulated that a biologist can artistically arrange a string of nucleotides[110] in a way that does not exist in nature,[111] there is a microscopically slim chance that such an arrangement could ever code for any protein or have any use other than as art.[112] This situation is unlike the phone directory in *Feist*, where subjective input and arrangement might have increased the ease of use and marketability.

Those arguing against the protection of databases must look to the future. The argument that the scientific community will be starved for information because everything that was in the public domain will be taken out is attenuated.[113] After all, it is possible that the new protection for databases will increase the production of useful databases. The researchers conducting basic research are not only database users but also database producers. If academic laboratories merge their resources and expertise, they can form comprehensive and useful databases. Naturally, the commercial sector will take notice and infuse money into academic research.[114] Academic laboratories could use this money to replace shrinking government funds, and to pay for additional research and educational activities.[115]

Gone are the days where academic labs generated all

---

able at http://www.bc.edu/bc_org/avp/law/st_org/iptf/articles/content/20000 12801.html

109. *Id.*

110. A nucleotide is "a compound consisting of a nitrogen-containing base linked to a sugar and a phosphate group. Nucleic acids are long chains of linked nucleotides, which in DNA contain the purine bases adenine and guanine and the pyramidines thymine and cystosine." "Nucleotide n." CONCISE MEDICAL DICTIONARY (Oxford University Press 2002).

111. Silva, *supra* note 108.

112. *Id.*

113. *See* Brudno, *supra* note 82, at 11.

114. "Over time. . .the distinction between "pure" or non-commercialized data and data applied to industrial pursuits seems likely to break down, as has routinely occurred in other disciplines. . .Universities and other research institutions may view data compilations generated in the course of research as potential revenue sources." J.H. Reichman & Pamela Samuelson, *Intellectual Property Rights in Data?*, 50 VAND. L. REV. 51, 68 (1997).

115. "Using data compilations as a source of revenue will be helpful to academic institutions, "especially in an era of declining government support for research endeavors." *Id.*

the basic scientific data. A lack of government funding for academia has propelled innovative and aggressive companies to fight for attention. Though the United States has been slow to respond to the growing need for database legislation, the European Union has enacted a *sui generis* regime to protect the investment of database compilers in the production of databases.

III. THE EUROPEAN UNION DATABASE DIRECTIVE—A RESPONSE TO *FEIST*.

The objective of the *sui generis* rights is not to promote the traditional goals of copyright law, but rather to protect the substantial investment of money and time in the development of databases.[116] Most countries' copyright laws do not extend protection to the contents of databases.[117] To the extent that database compilers cannot rely on encryption and contracts to prevent free riding, the European Union Database Directive (the "Directive") was implemented to close the gap in intellectual property law.[118] This *sui generis* regime[119] of property rights is meant to protect database compilers from free riding by second comers, and is in-

116. *See* European Union Database Directive, *available at* http://eon.law.harvard.edu/property00/alternatives/directive.html (last visited Feb. 14, 2003). For in-depth analysis of *sui generis* proposals, see G.M. Hunsucker, *The European Database Directive: Regional Stepping Stone to an International Model?*, 7 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 697 (1997). *See also* Reichman & Samuelson, *supra* note 114, at 51; Stephen M. Maurer et al., *Europe's Database Experiment*, SCIENCE, Oct. 26, 2001, at 789; Mark Schneider, *The European Union Database Directive*, 13 BERKELEY TECH. L.J. 551 (1998).

117. *See* Reichman & Samuelson, *supra* note 114, at 72.

118. *See id.*

119. The following justification for a *sui generis* regime has been given by the European Union:

> Even the mere accumulation of facts, statistics, bibliographical information, names and addresses involves considerable commercial activity. . .The data in this instance is similar to a raw material. If others misappropriate that raw material they will be able to market similar or identical products or services at greatly reduced costs. In other industries, it would be considered as an act of unfair competition for the raw material procured for processing at one company's expense to be freely appropriated by another company to make a similar product or service. On the other hand, no one manufacturer should have a monopoly over the source of the raw material such that he excludes others from the market for the finished product or service.

Mark Powell, *The European Union's Database Directive: An International Antidote to the Side Effects of Feist?*, 20 FORDHAM INT'L L.J. 1215, 1223 (1997).

tended to protect matters outside the scope of copyright law.[120]

Under the Directive, databases receive two levels of protection: "copyright protection for the original expression and arrangement of facts in the database. . .and *sui generis* protection prohibiting the unfair extraction of a substantial part of a database reflecting significant investment."[121] Neither right extends copyright protection to the underlying facts and data.[122] The European proposal required a "substantial investment in the compiling, verification, or presentation of information" for a database to be protected.[123] Creativity is not a prerequisite for protection under the new scheme.[124] If the database meets the "substantial investment" threshold, the database producers obtain the right to prevent extraction and re-utilization of the whole or a substantial part of the contents of the database.[125] The Directive allows users to extract or re-utilize "insubstantial parts" of the database so long as they do not harm the interests of the database producer.[126] "These new rights are granted on a reciprocal basis, which means that only authors from countries which grant similar rights to European databases are entitled to claim these rights in Europe."[127]

Commentators have argued that the Directive breaks away from the historical limits of intellectual property protection in at least three ways. First, it grants an exclusive property right based on the extent of investment without a requirement of creative contribution to the public domain. Second, it grants an exclusive property right in information

120. According to Powell, the European Union realized that copyright law alone would not provide the adequate level of protection for the substantial investments in purely factual databases. *Id.* at 1220.

121. Jonathan Band & Jonathan S. Gowdy, *Sui Generis Database Protection: Has Its Time Come?*, D. LIB. MAG., June 1997, *available at* http://www.dlib.org/dlib/june97/06band.html (last visited Feb. 17, 2003).

122. Hunsucker, *supra* note 116, at 741.

123. *See* Jane C. Ginsburg, *Copyright, Common Law and Sui Generis Protection of Databases in the United States and Abroad*, 66 U. CIN L. REV. 151, 171 (1997).

124. *See* Reichman & Samuelson, *supra* note 114, at 54-5.

125. *See* Hunsucker, *supra* note 116, at 723-24; Reichman & Samuelson, *supra* note 114, at 55.

126. *See id.* at 90-1.

127. Cynthia M. Bott, *Protection of Information Products: Balancing Commercial Reality and the Public Domain*, 67 U. CIN. L. REV. 237, 261 (1998).

that was previously considered an unprotectible raw material; and finally, it grants a perpetual right to the database owner as long as he continues to make a substantial investment in the database.[128]

These concerns must be mitigated by the incredibly rapid pace of change in electronic databases, which increase the rate of copying and re-distribution. As technology evolves, so must outdated methods of protection. It seems unfair that producers of factual compilations are left without protection for their investments simply because their business is in "facts." The Directive's two tiered approach effectively responds to this concern. The approach differentiates between compilations that truly are the result of original authorship and worthy of full copyright protection, and compilations that are not original but still worthy of protection because they are based on a large investment of labor and effort. Because this *sui generis* right is carefully balanced with appropriate fair use exemptions and restriction on duration, the public domain will be enriched by compilations of more useful databases rather than condensed because of information hoarding.

IV. LEGISLATION IN THE UNITED STATES

The database industry is part of a global information marketplace where development outside of the United States can have a marked effect on the ability of United States information providers to competitively operate internationally.[129] Databases are most valuable when they are

128. J.H. Reichman & Paul F. Uhlir, *Promoting Public Good Uses of Scientific Data: A Contractually Reconstructed Commons for Science and Innovation, available at* http://www.law.duke.edu/pd/papers/ ReichmanandUhlir.pdf. Commentators have argued that the European Directive is "one of the least balanced and most potentially anti-competitive intellectual property regimes ever created [and] will stiffly access to information, retard competition in the database industry, and impede basic scientific research." Hunsucker, *supra* note 116, at 706.

129. One commentator warns that the U.S. should not leap off the bridge like the E.U. It may turn out that American database producers can still compete with Europe. Additionally, trade sanctions can be imposed to deter the ill-effects of E.U. legislation on United States databases. *See* Maurer et al., *supra* note 116, at 789. However, the idea of sanctions against Europe is impractical. It would have a chilling effect on the advancement of research and harm the public in the long run. *See also*, Remarks of Q. Todd Dickinson, Under Secretary of Commerce for Intellectual Property and Director of U.S. PTO, *available*

comprehensive in their field of application, reliable, up-to-date, and logically organized. And these attributes can only be obtained through great labor and expense. But these are exactly the kinds of databases that are at risk because of the *Feist* decision. *Feist* and its progeny "reward artistic creativity, but not scientific utility."[130]

Federal copyright and state contract law still remain essential tools for protecting the large investments in databases. However, "there are clearly gaps that can be filled only by legislation for federal database protection."[131] Protection for non-creative database is necessary for three reasons: to establish an incentive for commercial producers to compile databases; to ensure accuracy and quality of the data; and to protect United States economic interests in a global information economy.[132]

The goal of Federal copyright legislation in the United States should be to reach a balance between providing incentives to private creators and maintaining sufficient public access to factual works.[133] The amount of work required to maintain, update, and verify databases is increasing exponentially and maximizing public resources.[134] However, the increased work is not matched by increased funding.[135] Federal government funding for scientific research is entirely limited by outdated budgets and numerous competing priorities and allocations. The possibility of government funding is significantly less likely considering the events of

---

at http:// www.uspto.gov/web/offices/ac/ahrpa/opa/bulletin/cardozo.pdf "I believe we need to move forward domestically . . . to ensure that the global intellectual property community does not move ahead without us." *Id.*

130. Statement of Dr. Robert Ledley before the Subcommittee on Courts and Intellectual Property, Oct. 23, 1997, *available at* http://www.house.gov/judiciary/41122.htm

131. Position Paper, Harvard Information Infrastructure Project, Advisory Committee on International Communications and Information Policy, *available at* http://www.ksg.harvard.edu/iip/acicip/IIA.HTM (last visited Mar. 8, 2003).

132. Michael Freno, *Database Protection: Resolving the United States Database Dilemma With an Eye Toward International Protection*, 34 CORNELL INT'L L.J. 165, 185 (2001). Freno's paper also provides a history of database protection in the United States. *Id.* at 168.

133. *But see* Glynn S. Lunney, Jr., *Reexamining Copyright's Incentives-Access Paradigm*, 49 VAND. L. REV. 483, 489 (1996) (Professor Lunney contends that the incentives-access paradigm is fundamentally flawed. He identifies allocative efficiency, rather than the incentives-access balance, as the appropriate guide for determining the proper scope of copyright protection).

134. *See* Reichman & Samuelson, *supra* note 114, at 145.

135. *Id.*

September 11 and the resulting War on Terrorism. Inevitably, public funds must be supplemented by private funds. However, "[w]ithout protection sufficient to justify investment in development and maintenance of updated collections, such investments will simply not be made; or will be made not by publishers of the information, but by those who would gain competitive advantage by keeping it to themselves."[136]

Some commentators have pointed out that after *Feist*, the growth rate of both databases and database producers dampened considerably in the following years.[137] This dampening demonstrates the increased need for *sui generis* protection rights. There are three main justification for this need: "Collections made through a substantial investment deserve protection from copying, [t]echnological change facilitates free-loading... [and] [n]ew rights for collection makers will improve their markets and rates of return."[138]

To address the recent legal developments, like the Directive, and technological developments, like electronic databases, intellectual property protection in the United States will have to be expanded.[139] If not, many of the American database industry's most valuable products face the grave prospect of being stolen and marketed as competing products.[140] Going beyond the traditional legal notions of protecting property rights, the United States has a concrete reason to enact legislation providing *sui generis* protection. This is because of the reciprocity requirement in the Directive. Unless the United States enacts complementary legislation, European companies can continue to extract valuable data from American databases while remaining protected from similar exploitation by American

136. *See* Statement of Dr. Robert Ledley, *supra* note 130.

137. *See* Freno, *supra* note 132, at 186.

138. Oram, *supra* note 54.

139. However, the expansion of intellectual property protection is not without its critics. *See* Deborah Tussey, *Owning the Law: Intellectual Property Rights in Primary Law*, 9 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 173, 233 (1998); J.H. Reichman & Paul F. Uhlir, *Database Protection at the Crossroads: Recent Developments and Their Impact on Science and Technology*, 14 BERKELEY TECH. L.J. 793, 812-13 (1999).

140. A previous attempt to enact domestic legislation was met with opposition from the scientific and engineering communities. *See* W. Gardner & J. Rosenbaum, *Database Protection and Access to Information*, SCIENCE, Aug. 7, 1998, at 786.

database companies.[141] This places American companies at a competitive disadvantage. In the growing electronic universe there is a greater need for a level playing field.[142]

The present success of the database industry nationally is an incorrect measuring stick for the industry's health at the international level.[143] Only a comparison between the United States and countries that protect their database through legislation could alleviate worries at the international level and demonstrate whether the United States is or is not lagging in production of useful databases.[144]

Further, the argument that the *sui generis* right will deplete the public domain and retard the creativity and growth of science is misconceived.[145] The added protection will not cause the demise of science because there are counter-balancing legal doctrines. Every proposal for added protection includes numerous exceptions that protect researchers and the public in certain contexts. Moreover, antitrust law can be used to address these and other problems.[146]

A. *The Collections of Information Antipiracy Act*

The Collections of Information Antipiracy Act (the Antipiracy Act)[147] was intended to "eliminat[e] the inequity in a legal regime that allows an unscrupulous competitor to copy with impunity the contents of someone else's compilation and then destroy the first compiler's market by selling a competing, less expensive product."[148] Opponents of database legislation[149] argued that the Antipiracy Act interfered

141. *See* Freno, *supra* note 132, at 189.

142. *See generally* Jack E. Brown, *Proposed International Protection of Electronic Databases*, 27 CUMBERLAND L. REV. 17 (1996).

143. Freno, *supra* note 132, at 189. *See also* Paula Baron, *Back to the Future: Learning from the Past in the Database Debate*, 62 OHIO ST. L.J. 879, 887 (2001) (Professor Baron proposes that a *sui generis* scheme based upon the principles of these cases, coupled with a shorter period of protection than that is available under copyright law, could provide the proper balance between access and incentive).

144. Freno *supra* note 132, at 189.

145. *Id.* at 190.

146. *Id.* at 191. *See generally*, A QUESTION OF BALANCE, *supra* note 74, at 8.

147. H.R. 354, 106th Cong. (1999). For a discussion of the Antipiracy Act, see Jonathan Band & Makoto Kono, *The Database Protection Debate in the 106th Congress*, 62 OHIO ST. L.J. 869 (2001).

148. Aaron Karnell, *The Collections of Information Antipiracy Act: Creating an Intellectual Property Right in Facts*, 28 S .U. L. REV. 1, 3 (2000).

149. Tussey, *supra* note 139, at 213-14.

with legitimate scientific research and reduced the amount of information available in the public domain.[150]

Under the Antipiracy Act, a database was considered a "collection of information," and was defined as "information that has been collected and has been organized for the purpose of bringing discrete items of information together in one place or through one source so that persons may access them."[151] Similar to the Directive, the Antipiracy Act prohibited a second-comer from extracting or utilizing a substantial part of the database if it would harm the actual or potential market of the primary user.[152] Whether a substantial part of the database has been taken depends on either a quantitative or qualitative analysis. The Directive, by way of contrast, only prohibits qualitative misappropriation. This analysis can be reduced to read that if the information is central to the databases worth, its extraction or reutilization is prohibited. This section essentially re-introduced the "sweat of the brow theory" because it protects investments of substantial monetary or other resources.[153] Any substantial investment would have been protected, regardless of whether it is money, time, or effort.

In addition, the Antipiracy Act, like the Directive, contained fair use exemptions to misappropriation similar to those enunciated in the Copyright Act of 1976.[154] For example, under the Antipiracy Act, non-profit, educational and scientific users would be able to extract or re-utilize data to the extent that their use did not harm the actual market of the product.[155] Section 1402, when read in conjunction with section 1403, would have extended the Antipiracy Act's pro-

150. *See* J. Ryan Mitchell, *If at Feist You Don't Succeed, Try, Try Again: An Evaluation of the Proposed Collections of Information Antipiracy Act*, 78 NEB. L. REV. 900, 916 (1999); Nelson, *supra* note 47, at 464.

151. H.R. 354 §1401 (1).

152. *Id.* at §1402.

153. Mitchell, *supra* note 150, at 911.

154. Copyright Act of 1976, 17 U.S.C. §107 (1976). Four factors in considering whether a particular use is a fair use are:

> the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; the nature of the copyrighted work; the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and the effect of the use upon the potential market for or value of the copyrighted work. *Id.*

155. H.R. 354 §1403(a)(1).

tection to a producer's potential market as well.[156] To keep from being too rigid, an exemption was carved out for "reasonable uses," provided that there was no direct harm to the producer's actual market.[157] As a result, in the genomics context, an academic researcher could not use a producer's database, make a discovery, and then sell that discovery to the original producer's competitor or compete with the original producer.[158]

Though some might argue to the contrary, *sui generis* rights will not take facts out of the public domain.[159] The Antipiracy Act contained a provision that would have allowed the extraction and re-utilization of insubstantial parts of a database or individual items of information.[160] Second users were not prohibited from independently gathering the same information through investment of substantial monetary or other resources. Protection under the Antipiracy Act was independent of copyright law and preempted state law remedies.[161] Though the debate which surrounded the Antipiracy Act continues today, the act itself is no longer a viable legislative option.[162]

B. *The Consumer and Investor Access to Information Act of 1999*

The Consumer and Investor Access to Information Act of 1999 (the Act of 1999)[163] defines a database as a "collection of discrete items of information that have been collected and organized in a single place, or in such a way as to be accessible thorough a single source."[164] The organization

156. *Id.* at §1402(a).

157. *Id.* at §1403(a)(1).

158. *See generally id.* at § 1403.

159. *Id.* at §1403 (c).

160. *Id.* at §1403 (b).

161. *Id.* at §1405(b), §1405(c). H.R. 354 also included a provision to federalize the "hot news" doctrine. *See* Boyarski, *supra* note 71, at 912 (1999). The "hot news" state misappropriation doctrine allows companies in media related industries to prevent their competitors from stealing time sensitive information such as current news items. *Id.* at 876. Boyarski contends that this provision gives much broader protection than the state misappropriation doctrine and therefore should not be adopted. *Id.* at 923.

162. *See* Blanke, *supra* note 17, at 680.

163. H.R. 1858, 106th Cong. (1999). *See generally* Band & Kono, *supra* note 147, at 873.

164. H.R. 1858 §101.

and collection must have required an investment of substantial monetary or other resources.[165] Similar to the Antipiracy Act, the Act of 1999 re-introduces the sweat of the brow doctrine.

Despite certain similarities, the Act of 1999 severely limits the rights of database producers against information pirates. For example, it only prohibits the distribution of duplicate databases in competition with the first database.[166] The distinction here places the task of determining the meaning of "duplicate" to the judiciary. Duplicate could mean 100% copying, or a "substantial similarity" in copying.[167]

The Act of 1999, like legislation preceding it, contains a fair use provision for scientific, research, or educational uses.[168] This time, however, the teeth have been pulled. The current legislation requires that liability will only occur if there is a consistent pattern of misappropriating the database for the purpose of direct commercial competition. The "consistent pattern" standard may be too arbitrary since harm to the primary database producer's market could be accomplished by one act of misappropriation. Despite this, a cause of action would only arise for the database producer if the person benefiting from the protection afforded under section 102 misused the protection.[169]

Two congressional terms have elapsed without a serious database protection bill. The introduction of a viable bill merits strong consideration. The Act of 1999 was a response to the criticism surrounding the harshness of the Antipiracy Act. However, the Act of 1999, in its present form, was not a serious bill for the protection of databases because it preempted all state law, and required a showing of duplication of the first database to establish infringement.[170] It seems that the United States legislature is

---

165. *Id.*

166. *Id.* at §102.

167. The phrase "substantial similarity" is derived from copyright cases involving computer programs and infringement. *See* Computer Assoc Int'l, Inc. v. Franklin Computer, Inc., 714 F.2d 1240 (3d Cir. 1983); Computer Assoc. Int'l v. Altai, Inc., 982 F.2d 693 (2d Cir. 1992).

168. H.R. 1858 §103(d).

169. *See* H.R. 1858 §102, §106.

170. H.R. 1858, "which is purportedly there to protect databases, really would dramatically reduce the protection we have for databases even today, without the bill." Symposium, *supra* note 18, at 278. Hansen also comments that the technical industry is generally contributing to the Republicans. *Id.* at

waiting for the European experiment to produce concrete results before stepping into the batter's box. Unfortunately, the time to act is now, and not when the ball has cleared the fence, leaving United States database producers and users facing an insurmountable deficit.

V. ALTERNATIVES TO A *SUI GENERIS* MODEL

Several commentators contend that, even in the absence of legislative protection, database publishers have alternative means of protection[171] through state misappropriation or contract law, and through technological self help measures such as encryption.[172]

The biggest flaw in using state law doctrines is the lack of national uniformity. Local protection is inadequate in light of the fact that databases are internationally accessible and marketable.[173] Furthermore, misappropriation doctrine is more useful in scenarios where the information is time-sensitive.[174] Human genome databases do not qualify because the sequences are fundamental information that will remain unchanged.

A. *State Contract Law*

State contract law is an efficient method of creating rights between two parties. Under the terms of the contract, database producers can customize the features of a database to the particular needs of users. Although contract law creates private rights on one hand, it lacks sufficiency and uniformity. The terms of the contract are only enforceable against a party to the contract.[175] Thus, once a party to the contract distributed the software to another party, the third party is not bound by the contract terms

299. Since both Houses are Republican, time for the legislation to pass has arrived.

171. Professor Ginsburg advocates a compulsory licensing scheme to strike a balance between the rights of database producers and users. *See* Jane C. Ginsburg, *Creation and Commercial Value: Copyright Protection of Works of Information*, 90 COLUM. L. REV. 1865 (1990)

172. *See* Terry M. Sanks, Comment, *Database Protection: National and International Attempts to Provide Legal Protection for Databases*, 25 FLA. ST. U. L. REV. 991 (1998).

173. *See* Barr, *supra* note 71, at 388-89.

174. *Id.*

175. *See* Barr, *supra* note 71, at 389.

because there was no privity of contract between the database producer and a third-party. This element of the "meeting of the minds" in not a requisite in copyright law.[176]

State contract law is not the appropriate method for increasing the grant of rights for database producers because it lack uniformity. This leaves open the possibility of conflicting results concerning the same database. Also, in an international, internet driven economy, the strength of a federal law is necessary to buttress the needs of database producers throughout the world.

A final shortcoming of state contract law is that it does not encompass the fair use provisions of copyright law.[177] Without the fair use provisions, educational institutions are at a disadvantage. The balance of rights is skewed. Database producers will not want to put their databases in an electronic format. Consequently, academic users will not have access to information for educational purposes.

State contract law is insufficient to meet the needs of database producers in an electronic economy. At a minimum, the protections of copyright law and a *sui generis* regime are required to ensure the continued production and distribution of databases.

B. *State Misappropriation Doctrine*

Under "state misappropriation doctrine,"[178] the court examines the methods which result in an unfair competitive advantage. These methods must consequently be prohibited.[179] This is measured by a balance of factors such as comparative costs, the value added, and the availability of similar copying or reverse engineering methods for other technologies.[180]

---

176. Similarly, state contract law does not preempt copyright law because copyright law does not share the bargained for exchange requirement. This extra element is not required to prove copyright infringement, therefore there is no preemption.

177. See

178. Professor Robert A. Gorman concluded that, in light of *Feist* and the preemption doctrine incorporated in section 301 of the Copyright Act, the Copyright Act preempts the state tort of misappropriation of facts and compilations. *See* Brown, *supra* note 142, at 24-25.

179. *See generally* Dennis S. Karjala, *Misappropriation as a Third Intellectual Property Paradigm*, 94 COLUM. L. REV. 2594 (1994).

180. Dennis S. Karjala, *Policy Considerations: Theoretical Foundations for the Protection of Computer Programs in Developing Countries*, 13 UCLA PAC.

The doctrine of state misappropriation also varies from state to state. Most recently, the Second Circuit, in *National Basketball Association v. Motorola, Inc.*,[181] enunciated a two-part test under New York's misappropriation doctrine. First, the Court looked to specific instances where the plaintiff generates highly time-sensitive information at some cost and defendant free-rides on plaintiff's efforts. Second, the Court considered whether the defendant competes directly with the plaintiff and whether defendants' actions would so reduce plaintiff's incentive to produce the service such that its existence or quality would be threatened.[182]

In some respects, state unfair competition law is inappropriate to address the lack of protection for the investment of effort in databases. To begin with, it is impractical to harmonize database protection schemes that exist in different forms in each state.[183] Secondly, the purpose of unfair competition law is to regulate the conduct of direct competitors, but not suppliers and users.[184] A legislative regime that "determines the acts to be performed without authorization by all users, whether or not they are competitors, is more desirable."[185]

Compilations of genetic information are unlikely to meet the tests for protection under state misappropriation law. Under the *Motorola* test, genomic publishers would not be able to show that their incentives are so severely reduced by copying as to threaten the business. Generally their profits are high and some mix of technological and contractual methods can be used to limit access to their information.

CONCLUSION

In the short term, protecting database makers from free-riders is the main priority because free-riding presents the greatest threat to the economic interests of database makers in the United States.[186] However, it is important to

---

BASIN L.J. 179, 188 (1994).

181. 105 F.3d 841 (2d Cir. 1997)

182. *Id.* at 845.

183. Powell, *supra* note 119, at 1224.

184. *Id.*

185. *Id.*

186. *Id.* at 1250.

realize that any regime for protecting databases must also be mindful of the future. For example, bioinformatics is a burgeoning area without specific metes and bounds. A delicate balance of promoting open access to data and incentives to produce data must be achieved. The pace of genetic research is largely dependent on the existence of comprehensive, and accurate databases. Unless private biotech companies are given an incentive to create databases we might not be far from imagining a slow crawl towards discovery and progress.

Factual compilations, such as databases, pose a particularly difficult problem for protection because their nature makes them difficult to adequately protect in the electronic age under the copyright law. Furthermore, creatively selecting data can exclude desirable information, while creatively arranging data may frustrate users.[187]

Legal scholars who continue to voice disapproval of databases based on the possible threat to social policy and the perceived risk of the removal of information from the public domain are misguided.[188] Their views are based on the false theory that intellectual property laws extend sweeping protection to the owners of patents, copyrights, and databases.[189] This is not true. Patents only protect specific inventions, not ideas. Copyrights only protect specific works, not types of literature or art.[190] Database rights would only protect specific data compilations, not raw factual information. These laws do not prevent someone from making the same compilation of data independent of the former work.

Database legislation is a matter of striking a balance. Researchers have to be encouraged to put data into the public domain rather than hoarding it as confidential information.[191] The interest of the scientific community at large also has to be considered. The grant of rights should be broad enough to promote investments in, rather than the stifling of research.[192] Database companies spend hundreds

187. *See* Hunsucker, *supra* note 116, at 716.

188. Polivy, *supra* note 77, at 777-78. Furthermore, the First Amendment and fair use exemptions adequately protect public rights. *Id.* at 778.

189. *See* Lavenue, *supra* note 10, at 36.

190. *Id.* at 58

191. Tim Powell, *Patenting the Human Genome*, *available at* http://www.eyeforpharma.com/index.asp?news=15978 (last visited Feb. 26, 2003).

192. *Id.*

of millions of dollars to collect, organize, verify, maintain and update huge databases. They also develop the best means to distribute reliable data to their customers quickly and efficiently, whether via the internet, in CD-ROMs or other formats.[193]

Currently, the copyright law does not protect the huge investments of time and money involved in database production and leaves the contents of these databases vulnerable to pirates. Copyright law does not adequately protect the investment of the producers. Federal legislation must be enacted to protect the investment of database compilers and to ensure that they remain a valuable and reliable resource for millions of users.[194]

193. *See* Coalition Against Database Piracy, *Separate the Facts from Fiction: The Truth About Database Protection Legislation*, *available at* http://www.cadp.net/legislation/legislationfacts.asp (last visited Mar. 17, 2003).

194. *See id.*