

ARTICLE

**DATABASE PROTECTION AT THE CROSSROADS:
RECENT DEVELOPMENTS AND THEIR IMPACT ON
SCIENCE AND TECHNOLOGY**

By J. H. Reichman and Paul F. Uhlir[†]

ABSTRACT

This article explores the potentially adverse impact that the emerging legal infrastructure could have on scientific, technical, and educational users of factual data and information—as well as on other sectors of the information economy—unless suitable adjustments are made. It begins by explaining how efforts to accommodate the networked environment to the publishers' fears of market failure will impose a daunting array of legal and contractual restraints on the ability of scientists and engineers to access factual data and information in the near future. It then goes on to examine the most recent efforts to devise a *sui generis* intellectual property right in noncopyrightable collections of data that would suitably balance public and private interests. It also emphasizes the need to reconcile legal protection of databases with fundamental constitutional mandates concerning free speech and the progress of science. The article concludes with a warning that overly protective initiatives could compromise the research-based institutions that currently ensure the technological predominance of U.S. industry in the global marketplace.

TABLE OF CONTENTS

I.	COMMODIFICATION OF DATA IN THE NETWORKED ENVIRONMENT: THE BIGGER PICTURE	796
II.	POTENTIAL IMPACT OF THE DATABASE PROTECTION LAWS ON SCIENCE AND TECHNOLOGY	799

© 1999 J.H. Reichman and Paul F. Uhlir.

† J.H. Reichman, Professor of Law, Vanderbilt University School of Law and Paul F. Uhlir, National Research Council, Washington, DC. An early version of this paper was presented to the Symposium on *The Changing Character, Use, and Protection of Intellectual Property*, German-American Academic Council in Cooperation with the U.S. National Academy of Sciences and the Max Planck Institute for Foreign and International Patent, Copyright and Competition Law, Washington, D.C., Dec. 3-4, 1998, and to the Conference on *Law in the Information Society*, Istituto per la Documentazione Giuridica, Comitato Nazionale della Ricerca, Florence, Italy, Dec. 2-5, 1998. The views expressed in this article are the authors' and not necessarily those of the National Academy of Sciences or the National Research Council.

A.	The User-Friendly Rules of Copyright Law	800
B.	Unbalanced Rules of the <i>Sui Generis</i> Model	802
1.	<i>Basic Substantive Principles</i>	803
2.	<i>The Resulting Legal and Practical Constraints</i>	806
C.	Long-term Implications of the <i>Sui Generis</i> Model	812
1.	<i>Reversing the Transparency Movement</i>	813
2.	<i>Transaction Costs Unlimited</i>	814
3.	<i>Endless Monopolies and Diminished Access to Government Data</i>	816
4.	<i>Gaming the Cooperative Ethos</i>	819
III.	RECENT DEVELOPMENTS: THE QUEST FOR AN APPROPRIATE UNFAIR COMPETITION APPROACH	821
A.	The Administration's Position	822
B.	A Negotiated Discussion Draft in the Senate	823
1.	<i>Clarifying the Demands on Scientific and Technical Users</i>	824
2.	<i>Compromise Proposals</i>	825
C.	Uncertain Future of the Database Protection Law	829
IV.	PRESERVING THE CONSTITUTIONAL BALANCE OF INTERESTS IN THE NETWORKED ENVIRONMENT	832
A.	The Competitive Ethos Under Attack	832
B.	The Constitutional Dilemma	833
C.	Erring on the Side of Caution	836

INTRODUCTION

The convergence of digital computing and telecommunications technologies has greatly expanded the already bright economic prospects for information goods of all kinds, but it has also unsettled the legal architecture on which the free market economies have previously been grounded.¹ Information products behave differently from the tangible, physical products of the Industrial Revolution;² and the legal paradigms that we have applied to balance incentives to create against both public good uses of

1. See, e.g., Julie E. Cohen, *Lochner in Cyberspace: The New Economic Orthodoxy of "Rights Management,"* 97 MICH. L. REV. 462 (1998); Trotter Hardy, *Property (and Copyright) in Cyberspace*, 1996 U. CHI. LEGAL F. 217 (1996); Robert P. Merges, *The End of Friction? Property Rights and Contract in the "Newtonian" World of On-line Commerce*, 12 BERKELEY TECH. L.J. 115 (1997); Henry H. Perritt, Jr., *Property and Innovation in the Global Information Infrastructure*, 1996 U. CHI. LEGAL F. 261 (1996).

2. See, e.g., Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources to Invention*, in THE RATE AND DIRECTION OF INVENTIVE ACTIVITY 609, 616 (National Bureau of Economic Research ed., 1962) (stressing that optimal utilization occurs when information is free, while optimal information production occurs only when producers expect to appropriate the economic value of their investments); see also Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CALIF. L. REV. 479, 591-608 (1995) (describing the economics of networks as "still under construction").

information and the discipline of free competition are stretched past the breaking point.³ We are thus challenged to rethink how best to structure competition for information goods in the emerging, worldwide information economy.⁴

The technological convergence that creates promising new markets for information goods also opens new opportunities for scientific and educational uses of data and information. However, a powerful movement to commodify data and information previously treated as a public good—that is, as an inexhaustible, indivisible, and ubiquitous component of the public domain⁵—could limit the ability of the scientific, technical, and educational communities to capitalize on such opportunities. The momentum generated by that movement would eventually have faced these communities with serious challenges even in the absence of a new intellectual property right in collections of data. The adoption of a strong property right in noncopyrightable collections of data by the European Union⁶—in a haphazard manner, with little serious economic or empirical investigation⁷—thus precipitated a crisis that was already well under way.

3. See generally, J.H. Reichman, *Legal Hybrids Between the Patent and Copyright Paradigms*, 94 COLUM. L. REV. 2432 (1994) [hereinafter Reichman, *Legal Hybrids*]; J.H. Reichman, *Charting the Collapse of the Patent-Copyright Dichotomy: Premises for a Restructured International Intellectual Property System*, 13 CARDOZO ARTS & ENT. L. J. 475 (1995) [hereinafter Reichman, *Charting*]; Pamela Samuelson, Randall Davis, Mitchell D. Kapor & J.H. Reichman, *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 COLUM. L. REV. 2308 (1994) [hereinafter Samuelson et al.].

4. See, e.g., Charles R. McManis, *Taking TRIPS on the Information Superhighway: International Intellectual Property Protection and Emerging Computer Technology*, 41 VILL. L. REV. 207 (1996).

5. See ROBERT COOTER & THOMAS ULEN, *LAW AND ECONOMICS* 40-41 (2nd ed. 1997) (noting that public goods are both non-excludable and non-rivalrous).

6. See 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. (L 77) 20 [hereinafter E.U. Directive].

7. See J.H. Reichman & Pamela Samuelson, *Intellectual Property Rights in Data?*, 50 VAND. L. REV. 51, 72-95 (1997) (tracing legislative history of the E.U. Directive). For different perspectives, see, for example, Robert C. Denicola, *Copyright in Collections of Fact: A Theory for the Protection of Nonfiction Literary Works*, 81 COLUM. L. REV. 516 (1981) and Jane C. Ginsburg, *Copyright, Common Law, and Sui Generis Protection of Databases in the United States and Abroad*, 66 U. CIN. L. REV. 151 (1997); 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 Wendy J. Gordon, *On Owning Information: Intellectual Property and the Restitutory Impulse*, 78 VA. L. REV. 149 (1992).

This article explores the potentially adverse impact that the emerging legal infrastructure could have on scientific, technical and educational users of factual data and information (as well as on other sectors of the information economy) unless suitable adjustments are made. Parts I and II explain how efforts to accommodate the networked environment to the publishers' and database makers' fears of market failure will impose a daunting array of legal and contractual restraints on the ability of scientists and engineers to access factual data and information in the near future. Part III examines the most recent efforts to devise a *sui generis* intellectual property right in noncopyrightable collections of data that would suitably balance public and private interests. Part IV emphasizes the need to reconcile legal protection of databases with fundamental constitutional mandates concerning free speech and the progress of science. It ends with a warning that overly protective initiatives could compromise the research-based institutions that currently ensure the technological predominance of U.S. industry in the global marketplace.

I. COMMODIFICATION OF DATA IN THE NETWORKED ENVIRONMENT: THE BIGGER PICTURE

Digital telecommunications networks enable publishers to control the uses of information goods directly by contract, without relying on state action to avoid market failure,⁸ for the first time since the advent of the Guttenberg printing press. In effect, online delivery has "restored the power of the two-party deal" with regard to information goods and diminished the dependence of publishers on artificial legal fences that copyright laws and other related rights supplied in the print environment.⁹

Efforts to accommodate the pre-existing legal landscape to the new technologies are proceeding along several different fronts. For example, because the new technologies empower publishers to fence off information goods by means of encryption devices and other technical protection

8. See, e.g., PAUL GOLDSTEIN, *COPYRIGHT'S HIGHWAY: FROM GUTTENBERG TO THE CELESTIAL JUKEBOX* 27 (1994) (stating that "[c]opyright was technology's child from the start [because] [t]here was no need for copyright before the printing press"); Wendy J. Gordon, *Asymmetric Market Failure and Prisoner's Dilemma in Intellectual Property*, 17 U. DAYTON L. REV. 853 (1992) (relating economic justification of intellectual property rights to the problem of market failure).

9. See J. H. Reichman & Jonathan A. Franklin, *Privately Legislated Intellectual Property Rights: Reconciling Freedom of Contract with Public Good Uses of Information*, 147 U. PA. L. REV. 875, 897-99 (1999) (discussing "The Restored Power of the Two-Party Deal").

measures,¹⁰ Congress has been persuaded to pass new laws making it a civil or criminal offense to disarm or tamper with these devices.¹¹ Would-be users must increasingly gain access to information goods via an electronic gateway where they are obliged to identify themselves and acknowledge the rights of the gatekeeper to the information goods, as, for example, expressed in copyright management information systems.¹² The new laws that defend the owner's encryption devices also forbid users from tampering with their intellectual property identity tags.¹³

At the same time, the National Commissioners on Uniform State Laws have proposed a model contract law to govern computerized information transactions that all state legislatures would eventually adopt. These proposals, until recently embodied in a draft Article 2B of the Uniform Commercial Code ("U.C.C."),¹⁴ would validate the publishers' standard form, non-negotiable contracts to which would-be users must assent in order to cross the electronic threshold and gain access to information de-

10. See, e.g., Julie E. Cohen, *A Right to Read Anonymously: A Closer Look at "Copyright Management" in Cyberspace*, 28 CONN. L. REV. 981, 983-89 (1996) (discussing technologies that copyright owners may utilize to monitor and control access to information); see also DanThu Thi Phan, Note, *Will Fair Use Function on the Internet?*, 98 COLUM. L. REV. 169, 192, 192 n.167 (1998) (defining and discussing "digital watermarks"). See generally NATIONAL RESEARCH COUNCIL, *CRYPTOGRAPHY'S ROLE IN SECURING THE INFORMATION SOCIETY* (Kenneth Dam & Herbert Lin eds., 1996); Mark Stefik, *Shifting the Possible: How Trusted Systems and Digital Property Rights Challenge Us to Rethink Digital Publishing*, 12 BERKELEY TECH. L. J. 137 (1997).

11. Digital Millennium Copyright Act, Pub. L. No. 105-304, §§ 103, 403, 112 Stat. 2860, 2863, 2889 (1998) (codified at 17 U.S.C. § 1201).

12. See Reichman & Franklin, *supra* note 9, at 897-98 (stating that this "gatekeeping function is reinforced by encryption devices, digital watermarking, and other self-help technical measures that permit information providers contractually to impose their own terms and conditions on access to information goods stored at any given network site and on the uses to which end-users can put the information they access").

13. See Digital Millennium Copyright Act, Pub. L. No. 105-304, § 103, 112 Stat. 2860, 2863 (1998) (codified at 17 U.S.C. § 1201).

14. See U.C.C. Article 2B—Licenses (Feb. 1999 Draft) (attempting to provide a common legal framework for transactions in digital information and software licenses). In April 1999, the American Law Institute, an original sponsor of this proposal, withdrew its support, and the National Commissioners announced their intention to pursue the project in the form of a model law governing computerized information transactions rather than as an amendment to the U.C.C. See NCCUSL & ALI, *NCCUSL to Promulgate Freestanding Uniform Computer Transactions Act: ALI and NCCUSL Announce that Legal Rules for Computer Information Will Not be Part of U.C.C.* (visited Apr. 23, 1999) <<http://www.nccusl.org/pressrel/2brel.html>>. For convenience, citations herein continue to refer to draft Article 2B, February 1999, the latest available version at the time of writing.

livered online.¹⁵ As matters stand, and despite mounting criticism from intellectual property scholars, the proposed model law would validate even "click on" or shrinkwrap licenses that ignored or attempted to override public interest exceptions that favored users or competitors, including the technical and scientific communities, under the pre-existing legal infrastructure.¹⁶ For example, such contracts could override the right to make non-infringing uses of copyrighted works or the right to reverse-engineer subpatentable innovation,¹⁷ and they could require payment for uses that courts had previously deemed fair uses under the federal copyright law.¹⁸

A third line of attack is to devise new intellectual property rights that, among other things, would serve to reduce potential tensions between state contract laws and the federal intellectual property system. As we shall see, copyright law expressly permits many uses of copyrighted works that publishers would like to restrict by means of online licenses.¹⁹ The validity of such contracts may be questioned on the grounds that they disrupt the federal intellectual property system (preemption arguments) or overstep the constitutional guarantees of free thought and expression.²⁰

However, if Congress enacted a hybrid ("*sui generis*") intellectual property right to protect the contents of databases, like that adopted by the European Union, it would give legislative approval to forms of protection that were previously unknown or questionable under traditional intellec-

15. Raymond T. Nimmer, the Reporter for Article 2B, justifies this approach. See Raymond T. Nimmer, *Breaking Barriers: The Relation Between Contract and Intellectual Property Law*, 13 BERKELEY TECH. L. J. 827 (1998).

16. See, e.g., Julie E. Cohen, *Copyright and the Jurisprudence of Self-Help*, 13 BERKELEY TECH. L.J. 1089 (1998); Jessica Litman, *The Tales that Article 2B Tells*, 13 BERKELEY TECH. L.J. 931 (1998); Charles R. McManis, *The Privatization (or "Shrink-Wrapping") of American Copyright Law*, 87 CALIF. L. REV. 173 (1999).

17. See, e.g., Reichman & Franklin, *supra* note 9, at 939-43, 947-51; see also Rochelle Cooper Dreyfuss, *Do You Want to Know a Trade Secret? How Article 2B Will Make Licensing Trade Secrets Easier (But Innovation More Difficult)*, 87 CALIF. L. REV. 191 (1999); Mark A. Lemley, *Beyond Preemption: The Law and Policy of Intellectual Property Licensing*, 87 CALIF. L. REV. 111 (1999).

18. See McManis, *supra* note 16, at 173 (addressing the capacity of Article 2B to alter the existing balance embodied in copyright law); Reichman & Franklin, *supra* note 9, at 943-47.

19. See 17 U.S.C. §§ 102(b) (1994) (ideas not protectable), 107 (1994) (fair use), 108-121 (1994) (other exceptions and limitations).

20. See 17 U.S.C. § 301 (1994); see also Dennis S. Karjala, *Federal Preemption of Shrinkwrap and On-line Licenses*, 22 U. DAYTON L. REV. 511 (1997); David A. Rice, *Public Goods, Private Contract and Public Policy: Federal Preemption of Software License Prohibitions Against Reverse Engineering*, 53 U. PITT. L. REV. 543 (1992).

tual property law.²¹ The creation of new intellectual property rights in collections of data would thus make it harder to resist arguments that publishers who subject online delivery of databases to technical protection measures and to contracts of adhesion that limit previously legal uses had violated fundamental public policies derived from the copyright laws. In other words, the database protection laws seem to permit acts (and foster policies) that overtly contradict or override the limits previously established by copyright and other traditional legal models.²²

Taken together, the ability of publishers to combine technical protection measures with tailor-made contract laws and hybrid intellectual property rights is supposed to stimulate investment in online commerce and to foster overall economic development.²³ Critics fear, however, that the cumulative effect of these separate but well-coordinated legal initiatives will be to balkanize the information economy and to unduly restrict the use of unbundled information as raw materials of science and technology or as inputs into the production of value-adding or second generation information goods.²⁴

II. POTENTIAL IMPACT OF THE DATABASE PROTECTION LAWS ON SCIENCE AND TECHNOLOGY

Let us suppose that a scientist or engineer lawfully obtained a printed copy of a chemical handbook or of a scientific article, with appended data, that was published in a peer-reviewed journal. These works currently at-

21. See E.U. Directive, *supra* note 6, arts. 7-10. For U.S. proposals spawned by the E. U. Directive, see Collections of Information Antipiracy Act, H.R. 354, 106th Cong. (1999); Collections of Information Antipiracy Act, H.R. 2652, 105th Cong. (1998); Database Investment and Intellectual Property Antipiracy Act of 1996, H.R. 3531, 104th Cong. (1996).

22. See Reichman & Samuelson, *supra* note 7, at 84-95, 103-110.

23. See, e.g., Nimmer, *supra* note 15; Hardy, *supra* note 1.

24. See *supra* notes 16-17; Reichman & Franklin, *supra* note 9, at 881 (stating that, to "ignore such discriminations as these is to risk watching model laws, adopted to govern the virtual marketplace for information goods, foster conditions that actually decrease innovation, discourage competition, and stifle the traditional marketplace of ideas"); see also G.E. Evans & B.F. Fitzgerald, *Information Transactions Under U.C.C. Article 2B: The Ascendancy of Freedom of Contract in the Digital Millenium?*, 21 U. NEW S. WALES L.J. 404 (1998) (arguing for the government's need to protect society given the recent shift of market power in the information economy).

tract copyright protection, and we shall assume that they meet the eligibility criteria of that body of law.²⁵

A. The User-Friendly Rules of Copyright Law

The rules of copyright law constitute a balanced regime of public and private interests. In retrospect, we are struck by the friendly treatment this body of law gives to users and competitors alike, notwithstanding the powerful bundle of exclusive rights it vests in authors and artists in order to stimulate the production and dissemination of creative works.²⁶

For example, any scientist or engineer who lawfully obtained the book or article mentioned above could immediately re-use all the data and all the ideas disclosed in them because copyright law does not protect ideas or data,²⁷ nor does it protect against use of expression as such, but only against certain specified uses.²⁸ Indeed, another scientist or engineer could independently rewrite his or her own version of the same article and disseminate it because copyright law allows independent creation, and all the unprotected data are spread out before the second comer's eyes.²⁹

A second scientist or engineer who needed to duplicate even the first author's creative selection and arrangement of data (if any) for non-profit research purposes could normally fall back upon the "fair use" provisions of current law.³⁰ A later researcher could also produce a follow-on article

25. See 17 U.S.C. §§ 101 (1994) (defining literary works), 102(a) (1994) (requiring original works of authorship), 103(b) (1994) (protection limited to original and expressive material added by author to a compilation); *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991) (limiting copyright protection of factual compilation to creative elements of selection, arrangement, and coordination).

26. See 17 U.S.C. § 106 (1994). See generally L. RAY PATTERSON & STANLEY W. LINDBERG, *THE NATURE OF COPYRIGHT: A LAW OF USERS' RIGHTS* 47-56, 191-224 (1991); JOEL SHELTON LAWRENCE & BERNARD TIMBERG, *FAIR USE AND FREE INQUIRY: COPYRIGHT LAW AND THE NEW MEDIA* (2d ed. 1989).

27. See *supra* note 25; *Harper & Row v. Nation Enterprises*, 471 U.S. 539 (1985) (stressing First Amendment interest in unrestricted availability of facts).

28. See 17 U.S.C. §§ 106, 106A (1994); *Baker v. Selden*, 101 U.S. 99 (1879); Ralph S. Brown, Jr., *Eligibility for Copyright Protection: A Search for Principled Standards*, 70 MINN. L. REV. 579, 588-89 (1985) (noting that the Copyright Act, unlike the Patent Act, does not confer any exclusive right to use the protected work, which helps explain why copyrights are so casually granted).

29. See, e.g., PAUL GOLDSTEIN, *COPYRIGHT* § 7.2.2 (2d ed. 1998) (noting that "conveying evidence" of independent creation constitutes a perfect defense to an infringement action).

30. See 17 U.S.C. § 107 (1994) (codifying fair use provisions); *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 564, 574-594 (1994) (stressing desirability of promoting transformative, preambular uses under the fair use provision of § 107); WILLIAM F.

or book that borrowed the originator's unprotected factual information and data, but not his or her stylistic expression.³¹ To be sure, the norms of world copyright law (but not necessarily U.S. law) favor attribution in such a case, as do the ethics of science.³² But plagiarism is not the same as copyright infringement; the reuse of facts and data is clearly permitted in copyright law; and another author's popularized version of a prior researcher's factual findings remains perfectly legal.³³

Most important, later scientists could combine the published data and factual information with other data and information into a multiple or complex interdisciplinary database without permission or additional payment to the originators.³⁴ This follows in part because only ineligible matter is at issue and in part because copyright law does not prohibit *use* as such, but only certain uses, such as reproduction or adaptation of protected expression, and it is also buttressed by the doctrine of fair use.³⁵

Even if scientists, engineers, or educators made classroom use of the protected expression for nonprofit purposes, these uses might well be fair

PATRY, *THE FAIR USE PRIVILEGE IN COPYRIGHT LAW* 178-84, 416-17 (1985). However, the wholesale duplication of a creative selection and arrangement for commercial purposes is not permitted. *See, e.g.,* CCC Info. Servs., Inc. v. Maclean Hunter Mkt. Reports, Inc., 44 F.3d 61 (2d Cir. 1994); Reichman & Samuelson, *supra* note 7, at 63 (citing authorities). *But see* Warren Publishing Inc. v. Microdos Data Corp., 115 F.3d 1509 (11th Cir. 1995) (en banc) (allowing massive extraction and re-use of commercial compilation for competitive purposes and finding no eligibility for copyright protection).

31. *See* 17 U.S.C. §§ 102(b), 103(b) (1994); Harper & Row v. Nation Enters., 471 U.S. 539 (1985); Key Publications, Inc. v. Chinatown Today Publ'g Enters., Inc., 945 F.2d 509, 514 (2d Cir. 1991) (stressing "thin" copyright protection doctrine of *Feist*). *See also* Reichman & Samuelson, *supra* note 7, at 63 (citing authorities).

32. *See* Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, as last revised at Paris, July 24, 1971, 828 U.N.T.S. 4, art. 6bis (obliging member states to respect moral rights of authors); Berne Convention Implementation Act of 1988, Pub. L. No. 100-568, 102 Stat. 2853 (1988) (implementing this obligation only indirectly).

33. *See* Hoehling v. Universal City Studios, Inc., 618 F.2d 972 (2d Cir. 1980); *supra* notes 31-32 and accompanying text. However, if a second comer were to pass off his article as that of the first scientist, grounds for invoking relief in unfair competition law might also exist. *See* Lanham Act § 43(a), 15 U.S.C. § 1125 (1998).

34. For the importance of this practice, in conjunction with the sharing ethos of science, see J. H. Reichman, *Why Science is Concerned About an Intellectual Property Right in Databases*, in AAAS SCIENCE AND TECHNOLOGY POLICY YEARBOOK 1998 (Albert H. Teich et al. eds., 1998), at 291, 301; *see also* International Council for Science (ICSU), Position Paper on Access to Databases, paper presented to the World Intellectual Property Organization (Sept. 1997) (unpublished manuscript, on file with authors).

35. *See supra* notes 26-31 and accompanying text.

or privileged uses under U.S. copyright law³⁶ and would possibly become subject to compulsory licensing under E.U. copyright laws.³⁷ Finally, having once purchased the book or the article, a scientist or engineer could sell it, lend it, or give it to others (first sale doctrine),³⁸ borrow it from a library,³⁹ use it as often as he or she liked for virtually any purpose, and make photocopies of it for scientific purposes under the fair use doctrine of U.S. law⁴⁰ or the private use doctrine of E.U. law.⁴¹

B. Unbalanced Rules of the *Sui Generis* Model

Now, let us suppose that the contents of the same chemical handbook or of the aforementioned scientific article were disseminated online and surrounded by technical fences as previously described. Suppose further that the contents of the book or article were protected by laws implementing the E.U.'s *sui generis* exclusive property right in noncopyrightable collections of data or by the U.S. version of that right, as set out in H.R. 2652, the Collections of Information Antipiracy Act (March 1998).⁴² The House of Representatives adopted H.R. 2652, and the Subcommittee on Courts and Intellectual Property then attached it to the Digital Millennium

36. See 17 U.S.C. § 107, 110(1) (1994); *but see* Princeton Univ. Press v. Michigan Document Servs., Inc., 99 F.3d 1381 (6th Cir. 1996) (en banc), *cert. denied*, 520 U.S. 1156 (1997) (limits on copies for classroom use).

37. See, e.g., Berne Convention, *supra* note 32, arts. 9(2) (limits on reproduction right), 10(2) (use for teaching purposes when consistent with fair practice); GUY TRITTON, INTELLECTUAL PROPERTY IN EUROPE 191 (1996). See also Lucie Guibault, General Report to the ALAI Annual Meeting (1998) (unpublished manuscript, on file with authors) (discussing exceptions and limitations in European copyright law).

38. See 17 U.S.C. § 109(a) (1994) (first sale doctrine).

39. See *id.*; see also 17 U.S.C. § 108 (1994) (reproduction by libraries and archives).

40. See 17 U.S.C. § 107 (1994); *supra* note 30. However, scientists and engineers working at for-profit institutions have lesser photocopying privileges, at least when secondary markets for photocopies and reprints are operational. See American Geophysical Union v. Texaco, Inc., 60 F.3d 913, 926-931 (2d. Cir. 1994); PATRY, *supra* note 30, at 190-94.

41. See European Commission Green Paper on Copyright and Related Rights in the Information Society, *reprinted in* 43 J. COPYRIGHT SOC'Y USA 50, 91 (1995) ("Most member states have introduced special legal arrangements for ... private copying..."). *But see* William R. Cornish, *Copyright in Scientific Works (Scientific Communications, Computer Software, Data Banks): An Introduction*, in EUROPEAN RESEARCH STRUCTURES—CHANGES AND CHALLENGES: THE ROLE AND FUNCTION OF INTELLECTUAL PROPERTY RIGHTS 50 (Max Planck Gesellschaft ed., 1994) (despite case for a measure of free reprography for purposes of academic research, "academic institutions are regarded as relatively soft targets by publishing interests [in U.K.], which ... [have been] inserting initial wedges.").

42. See E.U. Directive, *supra* note 6; H.R. 2652, 105th Cong. (1998).

Copyright Act, which became H.R. 2281, as sent to the Senate.⁴³ The database portion was dropped prior to Congressional enactment of that bill, however, and it was reintroduced with some modifications as H.R. 354 in January 1999.⁴⁴

1. *Basic Substantive Principles*

The *sui generis* provisions of the E.U. Directive⁴⁵ protect the contents of any noncopyrightable database that is the product of substantial investment against *extraction or reutilization* of the whole or of any substantial part (evaluated quantitatively or qualitatively).⁴⁶ Hence, this law could protect the noncopyrightable data appended to the hypothetical article in question or collected in the handbook, which the publishers might eventually disseminate online, with or without an accompanying print version.⁴⁷

Such protection lasts as long as new investments are made in updates or maintenance; hence perpetual protection of dynamic databases becomes a likely result, despite a nominal fifteen-year term.⁴⁸ There are no exceptions for "reutilization" by scientific and educational bodies, and there are no mandatory exceptions for "extraction" for scientific and educational purposes (although states may adopt this exception for noncommercial purposes).⁴⁹ The member states implementing the Directive must permit

43. See H.R. 2281, 105th Cong. (1998) (including the database protection bill as Title V).

44. See H.R. 354, 106th Cong. (1999).

45. See E.U. Directive, *supra* note 6, arts. 7-10. Besides a list of sixty "Recitals" or premises that underlie the legislation and a small set of definitional articles that apply across the board (arts. 1-2), the E.U. Directive also harmonizes the treatment of copyrightable databases in the member states' domestic laws. See *id.* arts. 3-6; Reichman & Samuelson, *supra* note 7, at 76-79. There is a final group of "common provisions" that apply to both copyrightable and noncopyrightable databases (arts. 12-16). While the provisions harmonizing the protection of copyrightable databases approximate the rules in the U.S., see *supra* notes 26-41 and accompanying text, they lie beyond the scope of this article.

46. See E.U. Directive, *supra* note 6, art. 7(1).

47. For the very broad definition of databases, see E.U. Directive, *supra* note 6, art. 1. For example, Reed Elsevier, Inc. has been buying up scientific journals and has recently begun to deliver scientific communications online.

48. See E.U. Directive, *supra* note 6, art. 10; Reichman & Samuelson, *supra* note 7, at 85-86.

49. See E.U. Directive, *supra* note 6, arts. 9, 9(b) (authorizing member states to allow extractions (but not reutilization) "for the purposes of illustration for teaching or scientific research, so long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved").

extraction or use of an insubstantial part of a protected database.⁵⁰ However, the risks of invoking even this exception are high, because a would-be user has no way of knowing in advance whether a court will later find that the amount used was in fact qualitatively or quantitatively insubstantial.⁵¹

U.S. bill H.R. 2652, later H.R. 2281 as part of the Digital Millennium Copyright Act, used different language to accomplish essentially the same result. It protected against *use or extraction* in commerce of all or a substantial part of a protected collection of information that is the product of substantial investment if such use or extraction would "cause harm to the actual or potential market" for a product or service that incorporated the collection.⁵² For this purpose, the term "collection of information" was very broadly defined,⁵³ and during face-to-face negotiation in the Senate (in which the authors of this article participated directly),⁵⁴ the publishers claimed that a single lost sale would fit within this standard of harm to the market. Any substantial new investment in updates or maintenance would prolong protection beyond fifteen years, with no limit to the number of renewals.⁵⁵ The bill initially recognized no exceptions for science and education as such, but, at the last minute, a provision tacked onto H.R. 2281 held scientists and educators liable only for harm to "actual mar-

50. See E.U. Directive, *supra* note 6, art. 8(1). Member states may allow a broader exception for "extraction for private purposes of the contents of a *non-electronic database*." *Id.* art. 9(a) (emphasis added).

51. See Reichman & Samuelson, *supra* note 7, at 90-91. See also *id.* at 87-95 (finding that the scope of protection under the E.U. Directive exceeds that of copyright law because 1) no idea-expression distinction is observed and no evolving public domain is generated, 2) the equivalent of a "derivative work" right in dynamic databases is not limited to new matter, and 3) the public interest exceptions are very narrow).

52. H.R. 2652, 105th Cong. § 1202 (1998); H.R. 2281, 105th Cong. § 1302 (1998).

53. See H.R. 2281, 105th Cong. § 1301(1) (1998) (defining "collection of information" to mean "information that has been collected and ... organized for the purpose of bringing discrete items of information together in one place or through one source so that users may access them"); *id.* § 1301(2) (defining "information" to mean "facts, data, works of authorship, or any other intangible material capable of being collected and organized in a systematic way"). See also *id.* § 1301(3) (defining "potential market" to mean "any market that a person claiming protection ... has current and demonstrable plans to exploit or that is commonly exploited by persons offering similar products or services incorporating collections of information").

54. See *infra* text accompanying notes 143-47.

55. See H.R. 2281, 105th Cong. § 1308(c) (1998). The interpretation in the text was confirmed by the position that the publishers took during face-to-face negotiations in the Senate. See *infra* text accompanying notes 143-47.

kets,” and not for harm to “potential markets” for their nonprofit uses of protected information.⁵⁶

The situation was further complicated in January, 1999, when Chairman Howard Coble of the House Committee on the Judiciary’s subcommittee dealing with intellectual property rights introduced a new version of the database protection bill, H.R. 354,⁵⁷ which modified the previous bill in at least two important respects. First, a serious effort was made to limit the term of protection to fifteen years, with little or no possibility of extension even in dynamic databases that are continuously updated.⁵⁸ Second, a new provision established an exemption for “additional, reasonable uses” by educational, scientific and research organizations,⁵⁹ which was loosely based on the “fair use” provisions of copyright law. However, this ambiguous provision would limit the proposed exception to “an individual act of use or extraction of information done for” specified purposes,⁶⁰ which apparently placed the burden of proof on the otherwise infringing researcher.⁶¹

Because one can read the new exception for scientific and educational uses set out in H.R. 354 broadly or narrowly, depending on how one interprets its latent ambiguities, it is instructive to assess the likely impact of the proposed legislation on science and technology as it stood at the end of 1998. We can then factor the proposed amendments into the analysis and compare them to certain promising proposals that emerged from face-to-face negotiations between stakeholders, held in the Senate under Senator

56. See H.R. 2281, 105th Cong. § 1303(d) (1998). An article attacking the proposed legislation that appeared in *Science* magazine focused particular attention on the harm to science from a “potential market” test. See William Gardner & Joseph Rosenbaum, *Database Protection and Access to Information*, 281 *SCIENCE* 786-87 (1998).

57. H.R. 354, 106th Cong. (1999). See also *Hearings on H.R. 354, the “Collections of Information Antipiracy Act” before the House Subcomm. on Courts and Intellectual Property of the House Comm. of the Judiciary*, 106th Cong. (1999) [hereinafter *Hearings*] (statement of the Honorable Howard Coble, Chair of the Subcommittee on Courts and Intellectual Property).

58. See H.R. 354, 106th Cong. § 1408(c) (1999) (limiting the term of protection to 15 years).

59. *Id.* § 1403(a)(2) (listing “Additional Reasonable Uses”).

60. *Id.* § 1403(a)(2)(A) (allowing individual act of use or extraction “for the purpose of illustration, explanation, example, comment, criticism, teaching, research, or analysis, in an amount appropriate and customary for that purpose ... if [such an act] is reasonable under the circumstances”) Criteria for determining reasonable use, and the limits thereon, are also set out. See *id.* § 1403(a)(2)(A)(i)-(iv).

61. See *infra* text accompanying notes 95-99.

Hatch's auspices in late summer of 1998.⁶² Accordingly, if Congress had adopted H.R. 2281 at the end of 1998, and that law had subsequently been applied to online delivery of the data contained in the book or article that were previously discussed in connection with the workings of copyright law,⁶³ the following results would have been likely to occur.

2. *The Resulting Legal and Practical Constraints*

In principle, a second scientist or engineer could not make any uses of the information or data that were not permitted by the form-contract site licenses that regulated access to the online database from which they were extracted.⁶⁴ The site license could charge one price for accessing or consulting the database, a second price for downloading it, and a third price for using it or reusing it in other contexts.⁶⁵

62. See *infra* text accompanying notes 140-65.

63. See *supra* text accompanying notes 26-41.

64. While this constraint could occur under existing law, the validity of such standard form contracts remains in doubt at the present time, with some courts upholding them and other courts invalidating them either on contracts grounds or under the doctrine of preemption. See Reichman & Franklin, *supra* note 9, at 876 n.1 (citing cases). The adoption of a model law governing computerized information transactions (like the previously proposed Article 2B of the U.C.C.), see *supra* note 14, would validate virtually all such contracts. See *id.* at 899-914 (criticizing this approach and proposing a new doctrine of "public interest unconscionability" to allow courts to reconcile freedom of contract with public-good uses of information). However, most opportunities to challenge the validity of such contracts as applied to either copyrightable or noncopyrightable databases on existing grounds would vanish if Congress adopted a database law along the lines of H.R. 2281, unless some other countervailing doctrine, such as the proposed "public interest unconscionability doctrine" became available. See Reichman & Franklin, *supra* note 9, at 947-51 ("Contracts Restricting the Use of Noncopyrightable Collections of Data").

If, instead, H.R. 354 were adopted, or a variant thereof that included a substantial exception for traditional scientific purposes, see *infra* text accompanying notes 148-52, the opportunities to challenge the validity of such contracts would depend on 1) the extent to which the database law itself restricted contractual overrides (none yet proposed, and all vigorously resisted by the publishers), and 2) the availability of ancillary doctrines in contracts law, such as the proposed "public interest unconscionability" doctrine. See *infra* text accompanying notes 163-64 (discussing the doctrine of misuse); see generally Reichman & Franklin, *supra* note 9, at 929-38 ("validating non-negotiable terms that respect the balance of public and private interests").

65. Cf. Reichman & Samuelson, *supra* note 7, at 117-24 ("Retarding the Progress of Science").

[T]he electronic publisher's growing capacity to charge for each and every use of online data (or at least for every "hit" that accesses such data), and to track and monitor every user ... means that it becomes in-

Even though the second scientist or engineer normally would have paid to access the data and information (and they are not copyrightable by definition), he or she could not use them in ways not permitted by the terms and conditions of that site license, which, in turn, would now be supported by a duly enacted federal intellectual property law.⁶⁶ Absent some constitutional override, the second comer could not, therefore, independently generate a similar article or study based on the same material without permission, even though the relevant data were now revealed to the public.⁶⁷ Because the data no longer entered the public domain,⁶⁸ he or she would need to obtain a new grant or substitute funding to repeat the collection process, in which case scarce funds would have been used to duplicate the creation of knowledge already in existence. This, of course, contradicts the norms of science, which favor building on previous discoveries and the sharing of research results.⁶⁹

In many instances, the data will be based on one-time events that later scientists and engineers could not physically regenerate, in order to fall within the permitted acts of independent creation under the database protection laws.⁷⁰ Even when regeneration remained feasible, the cost in rela-

creasingly capable of serving ... as its own collection society, subject to no consent decrees ... and no external regulation.

Id. at 153. At a recent conference on database protection in Italy that Professor Reichman attended, an Italian professor related that the European Commission had charged him a high price per page to consult official texts concerning antitrust laws and regulations and about double that price per page to download the same information for research purposes. See Tito Ballarino, Remarks at the University of Pavia Conference on "Le banche dati (anche su Internet)," Pavia, Italy, Oct. 2-3, 1997; cf. Jessica Litman, *The Exclusive Right to Read*, 13 CARDOZO ARTS & ENT. L. J. 29 (1994).

66. See *supra* note 64.

67. See H.R. 2281, 105th Cong. § 1303(b) (1998) (declining to "restrict any person from independently gathering ... or using information gathered ... by another person through the investment of substantial monetary or other resources"); *id.* § 1305(e) (allowing unrestricted licensing agreements). See also H.R. 354, 106th Cong. §§ 1403(c), 1405(e) (1999) (the same in this respect, but attenuated in impact owing to fair-use-like provisions set out in § 1403(a)). For possible constitutional overrides, see *infra* text accompanying notes 182-97.

68. Cf. Reichman & Samuelson, *supra* note 7, at 84-90 ("Abolishing the Public Domain"). Under H.R. 354, however, data would enter the public domain after fifteen years. See *supra* note 58.

69. See NATIONAL RESEARCH COUNCIL, BITS OF POWER: ISSUES IN GLOBAL ACCESS TO SCIENTIFIC DATA 1, 132 (1997) (stressing importance of scientific norms that favor the sharing of data and the cumulative process of acquiring scientific knowledge) [hereinafter BITS OF POWER].

70. See *supra* note 67.

tion to the niche market of likely users would normally be so high that few second comers would willingly regenerate the data.⁷¹ Hence, sole-source providers are likely to remain a dominant feature of the database landscape, real competition will continue to be the exception, and the strong property rights given database proprietors would potentiate existing barriers to entry.⁷²

Later scientists and engineers could not combine data legitimately accessed from one commercial database with data extracted from other databases to make a complex new database for addressing hard problems without obtaining additional licenses and permissions. This remains, perhaps, the single most critical problem for scientific and technical research.⁷³ Despite reassurances to the contrary from leaders of the international publishing community to leaders of the scientific community at a recent meeting in Paris,⁷⁴ lawyers representing publishers at face-to-face negotiations held in the Senate late in 1998 continued to insist that this customary and traditional scientific practice would, in principle, violate their redistribution rights.⁷⁵ Another critical factor is that there would never be a sale that exhausted the publisher's rights, only a license, which the proposed model laws of computerized information transactions would make perpetual.⁷⁶

No one could combine "substantial" amounts of data or information into a more efficient follow-on product without a license; the licensor

71. See, e.g., BITS OF POWER, *supra* note 69, at 114-24.

72. See *id.*; Reichman & Samuelson, *supra* note 7, at 90-95 ("Establishing Legal Barriers to Entry"), 124-30 ("Impeding Competition in the Market for Value-Adding Products and Services"). But see Laura D'Andrea Tyson & Edward F. Sherry, Statutory Protection for Databases, in *Hearing on H.R. 2652 Before the Subcomm. on Courts and Intellectual Property of the House Comm. on the Judiciary*, 105th Cong. (1997) (testimony of Laura D'Andrea Tyson, Consultant, Reed-Elsevier, Inc.) (contesting the strength of barriers to entry in database industry). New studies by the National Research Council later in 1999 will attempt to cast further light on these issues.

73. See ICSU Position Paper, *supra* note 34; *infra* notes 110-14 and accompanying text.

74. Trip report by Ferris Webster, Chair, ICSU/CODATA Group on Access to Data and Information, International Council for Science (1998) (on file with authors).

75. Whether the new exceptions proposed in H.R. 354 would alter this result remains to be seen. See *supra* notes 58-60, and accompanying text. See also *infra* text accompanying notes 140-46.

76. See U.C.C. § 2B-502 (Feb. 1999 Draft) (allowing prohibition of any transfer of mass-market licensed goods); Karjala, *supra* note 20, at 538; Reichman & Franklin, *supra* note 9, at 965. Unless otherwise restrained, such licenses could override the limited duration clause that H.R. 354 finally introduced. See *supra* note 58.

would labor under no duty to grant such a license: and the sole-source provider would not want any competition from follow-on products.⁷⁷ This also suggests, however, that the price would not be set so high as to encourage independent creation of the same data, when otherwise feasible. If so, and potential producers of follow-on products tended to invest in other activities, it would further discourage competition and innovation.⁷⁸

So long as natural and artificial barriers to entry remained high, scientists and engineers must pay artificially high prices to access commercial databases in the absence of competition. The enactment of strong exclusive property rights (complemented by strengthened contractual rights if the proposed model law were also adopted⁷⁹) thus seems likely to reinforce the pervasive sole-source character of the marketplace and exert further upward pressure on prices.⁸⁰

Meanwhile, scientists and engineers who paid to access protected databases could not routinely lessen overall transaction costs by lending, borrowing, or transferring the data they extracted to others working on a common problem. This follows because there would never be a sale or transfer under some equivalent of the "first sale" doctrine of the copyright (and patent) laws,⁸¹ only a license that would logically restrict further transfers without any time limit.⁸² Scientists and engineers who continued to share data once acquired without obtaining permission and without paying additional fees for such heretofore traditional or customary uses would "harm the market" that the database proprietor presumably secured by dint of the proposed legislation.⁸³

The data would not enter the public domain for at least fifteen years, and possibly never, if the private party were to continue to invest in main-

77. See Reichman & Samuelson, *supra* note 7, at 124-30. Cf. *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991) (sole-source proprietor of telephone directory denied permission to another directory compiler who wished to combine the data in the former's directory with data from numerous others whose owners had given their consent).

78. See *supra* notes 71-72.

79. See *supra* notes 14-15 and accompanying text.

80. Cf. BITS OF POWER, *supra* note 69, at 121-23 (chronicling failed attempt to privatize Landsat data in the 1980s, when prices of Landsat images rose from about \$400 per image to \$4,400 per image, a price at which the joint venture "was able to attract some commercial and federal customers, but few academic or independent researchers").

81. See *supra* notes 38, 76.

82. See Reichman & Franklin, *supra* note 9, at 964-65.

83. See H.R. 2281, 105th Cong. § 1302 (1998).

tenance or updates of a dynamic database.⁸⁴ Even data that nominally entered the public domain at expiry of the fifteen year term could remain unavailable in practice if would-be users lacked means to identify and isolate those data within the larger mix of protected and unprotected data comprising a dynamic collection.⁸⁵ If such data were rendered technically identifiable, nothing would prevent the proprietor from using electronic fencing devices and standard form contracts to further preclude extraction even after the intellectual property right had expired.⁸⁶

Moreover, unless proper precautions are taken, there is considerable risk that data generated or funded by the U.S. government would become privatized in ways that unduly restricted access on onerous terms and conditions.⁸⁷ If this were allowed to happen, taxpayer-financed data would be sold back to science and education at monopoly prices, with the likelihood that additional state subsidies would be needed to defray the costs. In the European Union, where governments intend to commercialize publicly funded data, insufficient thought has been given to this problem in general and to the impact on science and technology in particular.⁸⁸

A common thread uniting all the foregoing observations is the lack of any limits on the power of providers who benefit from legal protection of databases to impose any licensing terms or conditions they wish on access to, and use of, their products. In principle, the database provider could override by contract even the few exceptions and limitations contained in the bill, including the public's right to use insubstantial parts of a database.⁸⁹

84. *See id.* § 1308(c).

85. *See* Jane C. Ginsburg, U.S. Initiatives to Protect Works of Low Authorship, paper presented to New York University Conference on "Intellectual Products: Novel Claims to Protection and Their Boundaries," Engelberg Center on Innovation Law and Policy, La Pietra, Italy (June 25-28, 1998) (unpublished manuscript, on file with authors) (arguing that publishers should identify the expired components of protected compilations).

86. *See* Reichman & Franklin, *supra* note 9, at 897-913, 947-51.

87. *See, e.g., Hearings, supra* note 57 (statement of Andrew J. Pincus, General Counsel, U.S. Department of Commerce), at 13-20 (arguing for broad exemptions for government-funded data and warning about the "potential for "capture" of government-generated data) [hereinafter Statement of Pincus].

88. *See, e.g.,* Peter N. Weiss & Peter Backlund, *International Information Policy in Conflict: Open and Unrestricted Access versus Government Commercialization*, in *BORDERS IN CYBERSPACE* 300, 303 (Brian Kahin & Charles Nesson eds., 1997). *See also supra* note 65.

89. *See* H.R. 2281, 105th Cong. § 1303(a), 1305(e) (1998). *But cf.* E.U. Directive, *supra* note 6, art. 8(1). This E.U. privilege may not be overridden by contract, *see id.* art.

The net result, as Professors Reichman and Samuelson pointed out in an earlier article, is that, under the U.S. database proposals, as under the E.U. Directive,

the most borderline and suspect of all the objects of protection ever to enter the universe of intellectual property discourse—raw data, scientific or otherwise—paradoxically obtains the strongest scope of protection available from any intellectual property regime except, perhaps, for the classical patent paradigm itself.⁹⁰

When the provisions added to the latest bill, H.R. 354,⁹¹ are factored into the analysis, the end result is only slightly improved, at least in appearance if not in practice.

The first significant change mentioned above, which would more clearly inject protected data into the public domain after fifteen years,⁹² is of course a move in the right direction. However, the drafters still ignore the difficulties of identifying and accessing data whose term of protection had technically expired, an issue that was widely discussed last year.⁹³ The bill also ignores the power of database providers to override formal access to data that nominally entered the public domain by combining adhesion contracts with electronic fencing devices.⁹⁴

The second major change is a good faith effort to address some of the concerns of the scientific and educational communities by means of new, "fair-use-like" provisions.⁹⁵ However, these provisions are both ambiguous and too narrowly drawn.⁹⁶ By placing the burden of proof on scientists

15. United States publishers opposed the ban on contractual overrides of this provision in the E.U. Directive, and indicate that they intend to override it when permitted. See Mark Powell, *The European Union's Database Directive: An International Antidote to the Side Effects of Feist*, paper presented to the Fourth Annual Conference on "International Intellectual Property Law & Policy," Fordham University School of Law (Apr. 11-12, 1996) (unpublished manuscript, on file with authors).

90. Reichman & Samuelson, *supra* note 7 at 94.

91. H.R. 354, 106th Cong. (1999). See *supra* notes 58-61 and accompanying text.

92. See *id.* § 1408(c) (1999) (limiting the term of protection to 15 years).

93. See *supra* note 85 and accompanying text; Statement of Pincus, *supra* note 87, at 25-27 (questioning ability of users "to distinguish unprotected data entries from protected data entries" and fearing "*de facto* perpetual protection").

94. See H.R. 354, 106th Cong. § 1405(e) (1999) (allowing freedom of contract); Reichman & Franklin, *supra* note 9, at 899-914.

95. See H.R. 354, 106th Cong. § 1403(a) (1999).

96. See, e.g., *Hearings, supra* note 57 (statement of James G. Neal, Dean, University Libraries, Johns Hopkins University) ("[E]xemption for education and research ... remains far too narrow.") [hereinafter Statement of Neal]; *id.*, (testimony of Charles E.

and engineers, whose “individual acts” of “reasonable” use remain subject to scrutiny case by case,⁹⁷ they would continue to exert the chilling effect on research⁹⁸ that seems inherent in any “fair use” approach to a database law that does not otherwise provide the many other safeguards familiar from copyright law. Hence, as we explain below, a different kind of approach, one not strictly linked to the “fair use” concept, will be needed to ensure that a *sui generis* database regime does not harm customary and traditional scientific activities.⁹⁹

Even if a satisfactory legal formula to avoid harm to science and education were found, that formula would remain largely ineffective if database providers could simply override it by contract or, in the alternative, if the publishers could just charge more for access if they knew that the state would require them to charge less for extractions and reuse by scientific and educational bodies.¹⁰⁰ In short, unless the bill expressly and adequately immunizes traditional scientific and technical pursuits, the only limit on the database providers in most instances is what a monopoly market will bear.

C. Long-term Implications of the *Sui Generis* Model

We believe that the long-term implications of the proposed regime are potentially very damaging for science and technology. All science operates on databases. The near-complete digitization of data collection, manipulation, and dissemination over the past thirty years has ushered in what many regard as the transparency revolution.¹⁰¹ Every aspect of the

Phelps, Provost, University of Rochester, for AAU, ACE, and NASULGC) (“[E]xception for non-profit educational activities contains a broad, vague condition that vitiates its protection.”) [hereinafter Testimony of Phelps].

97. *See id.*

98. *See, e.g., Hearings, supra* note 57 (testimony of Joshua Lederberg, Nobel laureate, on behalf of NAS, NAE, IOM and American Association for the Advancement of Science (AAAS)) [hereinafter Testimony of Lederberg].

99. *See infra* text accompanying notes 151-55.

100. *See, e.g., Reichman & Franklin, supra* note 9, at 947-51; McManis, *supra* note 4.

101. *See, e.g., Paul F. Uhlir, From Spacecraft to Statecraft: The Role of Earth Observation Satellites in the Development and Verification of International Environmental Protection Agreements*, 2 GIS LAW 1 (1995). While we emphasize the impact of database protection on science and technology in this article, we predict that the larger economy will likewise suffer if the anticompetitive effects we foresee should materialize. There is hardly any sector of the economy that is not significantly engaged in the creation and exploitation of digital databases, and many—such as insurance, banking, or direct marketing—are completely database-driven. *See, e.g., Hearings, supra* note 57 (statement of

natural world, from the nano-scale to the macro-scale, all human activities, and indeed every life form, can now be observed and captured as an electronic database.

According to Nobel laureate Joshua Lederberg,

[d]ata are the building blocks of knowledge and the seeds of discovery. They challenge us to develop new concepts, theories, and models to make sense of the patterns we see in them. They provide the quantitative basis for testing and confirming theories and for translating new discoveries into useful applications for the benefit of society. They also are the foundation of sensible public policy in our democracy. 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.¹⁰²

1. Reversing the Transparency Movement

Science builds on science. In all areas of research, the collection of data sets is not an end in itself, but rather a means to an end, the first step in the creation of new information, knowledge, and understanding. As part of that process, the original databases are continually refined and recombined to create new databases and new insights. Typically, each level of processing adds value to an original (raw) data set by summarizing the original product, synthesizing a new product, or providing an interpretation of the original data.¹⁰³

The processing of data leads to a not readily apparent paradox. The original unprocessed, or minimally processed, data are usually the most difficult to understand or to use by anyone other than the expert primary user. With every successive level of processing, the data tend to become more understandable and frequently are better documented for the nonexpert user. As the data become more highly processed, documented, and formatted for easier use, they also are more likely to attract copyright protection.¹⁰⁴

Yet, it is the raw, noncopyrightable data that are typically of greatest use and value to researchers, who can manipulate and experiment with the

the Computer & Comm. Industry Assoc. of America and the Online Banking Association) (stressing how H.R. 354 threatens "legitimate reuse of information") [hereinafter Statement of CCIA et al.].

102. Testimony of Lederberg, *supra* note 98, at 5.

103. NATIONAL RESEARCH COUNCIL, PRESERVING SCIENTIFIC DATA ON OUR PHYSICAL UNIVERSE 16 (1995).

104. *Id.*; see also *supra* note 25.

original measurements in pursuit of their own research goals. If strong intellectual property protection of noncopyrightable data sets, which previously had the least commercial marketability, weakened the still nascent impetus toward transparency, it could disproportionately affect the availability of data most commonly used in basic research and higher education.

2. *Transaction Costs Unlimited*

The success of the U.S. basic research and educational system is predicated on the relatively unfettered access to and use of factual information; on a robust public domain for data; and on easy re-use, compilation, and value adding applications of data. Practically all databases developed in the pursuit of basic research and education are motivated by non-economic incentives such as the desire to create knowledge, the thrill of discovery, and the enhancement of professional status.¹⁰⁵ The new database laws, however, place an overriding emphasis on protecting original investments and on augmenting purportedly necessary economic incentives to create new databases. At the same time, they undervalue the adverse effects on scientific and technical progress, as well as the aggregate economic and social costs inherent in restricting and discouraging the downstream applications and transformative uses of noncopyrightable databases in general.¹⁰⁶

The lack of any restraints on licensing, especially on sole-source data providers, adds to the dangers inherent in the creation of a strong exclusive property right in collections of data.¹⁰⁷ In particular, the ability of data providers to override by contract even the limited exceptions that the new law may grant to public-interest users, including scientists, engineers, and educators, is of great concern. Without a concomitant duty to deal fairly and reasonably with public-interest users, these combined powers could lead to high prices for data and to the imposition of harsh and oppressive

105. *Cf., e.g.*, BITS OF POWER, *supra* note 69, at 113 (comparing contributors and users of scientific data to non-market models of "a family or clan, in which exchange is not monetized but depends on social norms specifying expected and well-understood levels of contribution").

106. *See, e.g.*, Testimony of Phelps, *supra* note 96, at 3-4 ("The Academic Environment and Activities threatened by H.R. 354").

107. *See, e.g.*, Ginsburg, *supra* note 7, at 175 ("When the data ... [are] not available elsewhere ... the potential breadth of the potential market is very troublesome.").

terms concerning both access and subsequent uses of data that would especially disadvantage academic researchers.¹⁰⁸

Moreover, scientists and engineers will have to defray increased transactional and administrative costs engendered by the need to enforce the different legal restrictions on newly obtained data, to institute new administrative guidelines regulating institutional acquisitions and uses of such data, and by associated legal fees. Because universities and government agencies are inherently conservative, risk-averse institutions, they will err on the side of caution and place additional limits on what researchers and educators can do in acquiring and using data in order to avoid the possibility of costly litigation.¹⁰⁹

The proposed database law would severely discourage the re-use, re-compilation, and other value adding uses of data. Anytime someone uses data in a "collection of information" protected by the proposed law, that user becomes exposed to claims that he or she will have harmed the database originator's actual or potential markets.¹¹⁰ As a practical matter, this means that once public-domain data are collected and used for one purpose, such as to prepare a compilation of poisons and antidotes, it will foster a strong disincentive to use the same data for other purposes lest those uses violate the "harm to other markets" principle. By the same token, database recompilers or value adders incur the risk of lawsuits for infringement every time their new database resembles some pre-existing database, whether those data were used or not.¹¹¹

One of the most serious problems of all is the risk of inhibiting the creation and exploitation of multiple-source data products, which have be-

108. See, e.g., Testimony of Lederberg, *supra* note 98, at 2-6 ("Progress in the creation and reuse of new knowledge for the national good depends on the full and open availability of government and government-funded data, and on fair and equitable availability of data from the private sector."); Reichman & Franklin, *supra* note 9, at 943-47.

109. For the problems that university administrators already foresee, see Testimony of Phelps, *supra* note 96.

110. See H.R. 354, 106th Cong. § 1402 (1999). However, unauthorized extractions or uses for *nonprofit* educational, scientific, or research purposes incur liability only for harm to actual markets. See *id.* § 1403(a)(1).

111. See, e.g., Statement of CCIA et al., *supra* note 101; Ginsburg, *supra* note 85, at 23-24 (stressing need for publisher to identify value-added contributions). The exception that permits anyone to make use of "insubstantial parts" of a collection of information is vitiated by the language inflicting liability for harm to the investor's "actual or potential market." See *supra* notes 53-55 and accompanying text. Because the user cannot know such matters in advance, the "potential harm" test emasculates the "insubstantial parts" exception in practice.

come the scientific method of choice for addressing hard new problems. Because research is increasingly conducted by teams, often operating from different institutions, the pertinent data “are drawn from multiple sources, recombined and merged with new data to produce data sets that may lead to new and unanticipated findings.”¹¹² As Joshua Lederberg testified at a hearing on H.R. 354, the “recent advent of digital technologies for collecting, processing, storing, and transmitting data has led to an exponential increase in the size and number of databases created and used. A hallmark trait of modern research is to obtain and use dozens or even hundreds of databases, extracting and merging portions of each to create new databases and new sources for knowledge and innovation.”¹¹³

In this regard, the Administration itself predicted that, under the current proposals, scientists and engineers would face rising transaction costs when attempting to create complex databases from multiple public and private sources. Also predicted are higher costs due to the burdens of administering national data centers and of carrying out related, large-scale management activities that currently benefit from the policy of open and unrestricted access to scientific and technical data.¹¹⁴

3. *Endless Monopolies and Diminished Access to Government Data*

Because many data providers are sole-source and an exclusive property right would greatly strengthen the legal and economic protection of these mini-monopolies, the proposed legislation seems likely to raise the costs of data acquisitions to researchers and educators generally, not to mention other consumers. Those costs would either be passed on to the government and the taxpayer through increased research contract and grant requests, or they would simply diminish the resources available to researchers and education. If the costs and restrictions on all downstream or transformative data users—whether in the public or private sector—similarly increased (as feared by database proprietors opposed to strong

112. Testimony of Phelps, *supra* note 96, at 3.

113. Testimony of Lederberg, *supra* note 98, at 6. See also Testimony of Phelps, *supra* note 96, at 3 (“In the academic community, ... databases are dynamic instruments: they are not only sources of information, but they themselves—or components of them—become ingredients in new products, both through the combination of multiple contemporaneous data sets to produce qualitatively new products, and through the re-analysis of prior data from new perspectives provided by new findings or new analytic tools.”).

114. See Letter from Andrew J. Pincus, General Counsel of the U.S. Dept. of Commerce, to Senator Patrick J. Leahy, Ranking Minority Member of the Senate Committee on the Judiciary (Aug. 4, 1998) (on file with authors) [hereinafter Administration’s Letter 1998].

ilarly increased (as feared by database proprietors opposed to strong protection), it would discourage socially and economically beneficial forms of exploiting factual data that have up to now been available from the public domain.¹¹⁵

The fifteen year term (which is potentially much longer because of loopholes favoring constantly updated, dynamic databases) is particularly likely to hamper the progress of science and technology, a prospect that troubles the Administration, too.¹¹⁶ Such long delays in unfettered access to and use of data will undermine the value of many data sets for most fields of research, including research pertaining to the formulation of government policy; and in other cases it will effectively remove them from comparative analysis with other, openly available, concurrent data sets.¹¹⁷

A fifteen year period appears completely arbitrary and has not been seriously compared with other, potentially shorter, periods of protection.¹¹⁸ The proposed legislation thus defeats a primary, constitutionally mandated purpose of intellectual property laws, which is to establish a public domain that "promotes science and the useful arts,"¹¹⁹ from which researchers, educators, and other downstream users can build on previous contributions to further knowledge.

The proponents of the legislation say that nothing prevents a user or competitor from independently creating an equivalent database.¹²⁰ But many databases cannot be recreated from scratch. Data that are time-sensitive, unique, very old, or prohibitively expensive fit this description. In research, this includes virtually all observational data sets of transient natural phenomena, as well as data from very costly or labor-intensive experiments. Furthermore, a basic underlying principle in research and edu-

115. See Statement of CCIA et al., *supra* note 101, at 8-10 (noting that some commercial database providers oppose strong protection owing in part to fears about consequences of sole-source market structure).

116. See *supra* note 55 and accompanying text; Statement of Pincus, *supra* note 87, at 21-27.

117. Besides retarding scientific and technical progress, the fifteen-year term has no apparent justification in the rapidly moving commercial database industry either, where economic exploitation of most data products is typically measured in months and years, and even minutes and hours, rather than decades. See Statement of CCIA et al., *supra* note 101, at 10.

118. See, e.g., Ginsburg, *supra* note 85, at 30 (acknowledging soundness of lead-time criterion not adopted in E.U. Directive); Reichman & Samuelson, *supra* note 7, at 145 (proposing lead-time criterion and relating it to lack of actual or legal secrecy).

119. U.S. CONST. art. I, § 8, cl. 8.

120. See, e.g., Ginsburg, *supra* note 7, at 175.

ation is that the creation of new knowledge should build on the base of existing data and information, and that scientists and engineers should not have to duplicate previous factual compilations or discoveries in socially and economically inefficient ways.¹²¹ Protection of investments in factual databases is not the only interest that the law should seek to protect in this area.

Public interest users in the United States are likewise concerned about the applicability of the E.U. Database Directive to government data and the potential restrictions on access to and use of European public-sector data. Moreover, even though the proposed U.S. legislation does expressly exempt government data from its scope of protection, there are concerns that, as drafted, this exemption could be circumvented in several ways.¹²²

This can occur, for example, if the contractors or grantees are not expressly required either to provide their data back to the government for public dissemination, or to make the data publicly available themselves under appropriate terms and conditions.¹²³ Absent such universal vigilance by the government, a lot of data produced as a direct result of public funding could end up under proprietary control of researchers or their institutions. Because most of the noncopyrightable databases generated with government funding in the United States are actually created by non-government employees, whether in academia or industry, the failure of government agencies to enforce this exemption could have a far-reaching impact on the full and open availability of publicly funded data. Indeed, there is some risk that government agencies could increasingly view database protection as an income-generating opportunity, like their European

121. See, e.g., Statement of Lederberg, *supra* note 98, at 4-5; Testimony of Phelps, *supra* note 96, at 3-4; Statement of Neal, *supra* note 96.

122. See H.R. 354, 106th Cong. § 1404 (1999); see also, Statement of Pincus, *supra* note 87, at 13-20 ("Third Principle—Preserve access to government data"). The federal basic research budget alone is estimated to be over \$19.5B/year, of which a sizeable fraction is devoted to the creation, maintenance, dissemination, and analysis of scientific and technical data. See INTERSOCIETY WORKING GROUP, AMERICAN ASS'N FOR THE ADVANCEMENT OF SCIENCE REPORT XXIV: RESEARCH AND DEVELOPMENT FY 2000 71 (1999). Moreover, the government at all levels produces data of other importance to the nation, including economic growth, public health and safety, regulatory requirements, cultural affairs, and many other functions. All citizens have an interest in preserving full and open access to all government data that are not otherwise restricted by national security, privacy, or other legitimate limitations.

123. See, e.g., Ginsburg, *supra* note 85, at 28 ("In the case of sole-source government information, public domain policy favors making the information available to market rivals.").

counterparts.¹²⁴ As more university research is funded by private sources, more data will likely be removed from the public domain in the form of income-producing products.

Still other legislation, if combined with increased database protection, could further limit the principle of full and open access to government data. For example, the Commercial Space Act of 1998 encourages NASA to purchase space and earth science data collection and dissemination services from the private sector and to treat data as commercial commodities under federal procurement regulations.¹²⁵ When coupled with strong protectionistic measures, such as those contemplated by H.R. 354, we could eventually witness the passing of substantial amounts of data from the public domain of entire federal agencies. It also remains unclear if the government concludes an arrangement with a private sector party to disseminate public data or information, whether there will be adequate safeguards that either promote competition or that require low-cost access for public-interest users.¹²⁶

4. *Gaming the Cooperative Ethos*

Finally, a high-protectionist regime tends to undermine scientific and technical cooperation over time and to exert a progressive chilling effect on data-intensive research. As scientists, engineers, and their employing institutions become more accustomed to a new legal regime that encourages the commercial exploitation of their own research data sets, the cooperative culture that has become the hallmark of so many fields of science will be threatened.¹²⁷ Universities have already indicated that they intend to commercially exploit databases, and they have obtained an exemption for state universities from the government data exception in the proposed legislation.¹²⁸ If scientific institutions in one segment of the research community try to commercially exploit their colleagues in other institutions or countries, still others will be tempted either to emulate such be-

124. See *supra* note 88 and accompanying text.

125. See Commercial Space Act of 1998, Pub. L. No. 105-303 (1998).

126. Cf. Statement of Pincus, *supra* note 87; Ginsburg, *supra* note 85, at 27-29.

127. See, e.g., BITS OF POWER, *supra* note 69, at 169-70 (urging the scientific community to organize its own administration of data in order to preserve the sharing ethos).

128. See H.R. 354, 106th Cong. § 1404(a)(1) (1999) (limiting the term of protection); Testimony of Phelps, *supra* note 96, at 17 (noting that "universities and colleges are not only users of compilations of information, they also act as creators of collections that should be [entitled to] protection to the same extent as collections created by commercial providers").

havior or to cut off cooperation. Either way, science and technology would suffer.

Even if scientific data exchanges in established cooperative research programs were allowed to continue among a select group of principal investigators and an approved class of associated researchers, it would become increasingly difficult for other researchers outside the officially sanctioned group to obtain full and open access to the program data. This result, of course, would discourage interdisciplinary research and applications, contrary to the interests of technological innovation and the advancement of knowledge.

If simple exchanges of data and access to single databases became legally threatening or prohibitively expensive, imagine the potential transactional burdens that ill-conceived laws could impose on data compilers or users who needed to integrate data from multiple, or even hundreds, of different sources. This brings us to what may well be the most profound—and insidious—impact of the proposed legal regime on science and technology: the lost opportunity costs that will be repeated thousands of times each day across the basic and applied research communities. If scientists and engineers must choose between spending a lot of administrative time and a larger percentage of their valuable research grants on acquiring data and doing other, less data-intensive work, they will increasingly opt for the second route, despite the astounding yields that have so far been harvested from data-intensive research under existing conditions.

For all these reasons, an overly protective database regime would seriously impede the use, reuse, and transformation of the factual data that are the lifeblood of science and technology. In a worst case scenario, this law would first disrupt the system of cheap access to upstream data for purposes of basic research, it would then lead to ever higher prices for the acquisition of data used in applied research, and finally, it would strangle the ability of value-adding researchers and industries to improve, transform, or develop follow-on databases and related information products.¹²⁹ These outcomes would, in turn, greatly reduce the downstream applications of scientific breakthroughs subject to exclusive property rights. In sum, putting a strong property right too far upstream too soon¹³⁰ could have a disastrous effect on the long-term competitiveness of the U.S. economy and

129. See, e.g., Statement of CCIA et al., *supra* note 101.

130. Cf. Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCIENCE 698-701 (1998).

would undermine a key comparative advantage this country enjoys in the high-tech sectors of the global marketplace.

III. RECENT DEVELOPMENTS: THE QUEST FOR AN APPROPRIATE UNFAIR COMPETITION APPROACH

Concern about these issues mobilized the scientific, educational and library communities to express their views both to Congress and to the World Intellectual Property Organization ("WIPO"). For example, the Presidents of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine sent several letters to the Administration and to leading members of Congress responsible for this legislation.¹³¹ The International Council for Science ("ICSU") likewise intervened at relevant meetings of WIPO, and documents submitted by ICSU have played a prominent role in successful efforts to block rapid or premature efforts to launch an international treaty regulating databases modeled on the unbalanced E.U. Directive.¹³² ICSU has also begun direct consultation with publishers' representatives, with a view to working out some common understanding applicable to database protection issues affecting the scientific community.

Initial efforts in 1997 and early 1998 to slow the legislative process in the U.S. House of Representatives were not successful. The concerns of the scientific community were largely ignored by the House Committee on the Judiciary's Subcommittee on Courts and Intellectual Property, which first pushed H.R. 2652 (the "Collections of Information Antipiracy Act") through the House in the spring of 1998 and then had that bill attached to the Digital Millennium Copyright Act, H.R. 2281, in July 1998.¹³³

131. Letters from the three Academy Presidents to Mickey Kantor, Secretary of Commerce (Oct. 9, 1996) and to Senator Orrin Hatch, Chair of the Senate Committee on the Judiciary (July 10, 1998) (on file with authors).

132. See ICSU Position Paper, *supra* note 34; Reichman & Samuelson, *supra* note 7, at 95-102, 110-13 (describing efforts to have WIPO adopt a database protection treaty in 1996).

133. See H.R. 2652, 105th Cong. (1998); H.R. 2281, 105th Cong. (1998); see also COPYRIGHT OFFICE, REPORT ON LEGAL PROTECTION FOR DATABASES (1997), available at <<http://lcweb.loc.gov/copyright/reports>>. For the latest position of the Copyright Office, see *Hearings*, *supra* note 57 (statement of Marybeth Peters, Register of Copyrights) (noting improvements in H.R. 354, but stating that "it is important not to inhibit or raise the cost of public interest uses" and that an "appropriate statutory balance should result in optimizing the availability of reliable information to the public") [hereinafter Statement of Peters].

A. The Administration's Position

More recent developments, however, have been favorable to the interests of the scientific and educational communities. To begin with, an interagency review initiated within the Administration in the spring of 1998 produced a series of important position papers that supported the theses that the scientific community had already put forward. On August 4, 1998, the General Counsel of the U.S. Department of Commerce, Andrew Pincus, wrote Senator Patrick Leahy, Ranking Minority Member of the Senate Committee on the Judiciary, to advise that any new database legislation must avoid capture by private parties of government data and that "any effects [it may have] on non-commercial research should be *de minimis*."¹³⁴ This letter was sent as a consensus position of all departments and agencies of the Administration.

Consistent with these views, the Administration expressed concerns about possible

increase[d] transaction costs in data use, particularly where larger collections integrate data sets originating from different parties or where different parties have added value to a collection through separate contributions.... This is especially important for large-scale data management activities, where public investment has leveraged contributions from the private and non-profit sectors.¹³⁵

The letter went on to express further concerns "that the ... exception for noncommercial research and educational uses does not ensure that legitimate non-commercial research and educational activities are not disrupted by the prohibition against commercial misappropriation" and that sole-source providers might unduly burden "access and use" by this sector.¹³⁶ The Administration reiterated most of these same concerns in testimony before the House Subcommittee on Courts and Intellectual Property, at hearings concerning H.R. 354 on March 18, 1999.¹³⁷

The Administration's initial letter of August 4, 1998, also referenced the Department of Justice's "serious constitutional concerns that the First Amendment restricts Congress's ability to enact legislation" of this kind, and that other constitutional obstacles would have to be overcome.¹³⁸

134. Administration's Letter 1998, *supra* note 114.

135. *Id.*

136. *Id.*

137. See Statement of Pincus, *supra* note 87.

138. Administration's Letter 1998, *supra* note 114.

These constitutional impediments were elaborated in a 26-page memorandum by the Legal Counsel of the U.S. Department of Justice, dated July 28, 1998, which detailed a serious indictment of the *sui generis* model then pending before Congress.¹³⁹

On September 28, 1998, moreover, the Chair of the Federal Trade Commission ("FTC"), Robert Pitofsky, wrote the Chair of the House Committee on Commerce, Tom Bliley, to express additional concerns about the pending database legislation.¹⁴⁰ In particular, the FTC found that "certain provisions within the proposed legislation raise concerns about possible unintended, deleterious effects on competition and innovation," and that "the potential for anti-competitive use of a 'collection of information' is substantially increased when there is only a single source for the data."¹⁴¹

Finally, it should be noted that the United States Patent and Trademark Office ("USPTO") held a conference on April 28, 1998, to reexamine database protection and access issues. In July 1998, it issued a Report that, while endorsing a modified form of *sui generis* database protection, expressed support for some of the concerns that the scientific and educational communities had been voicing.¹⁴²

B. A Negotiated Discussion Draft in the Senate

In late July of 1998, the Chair of the Senate Committee on the Judiciary, Senator Orrin Hatch, invited representatives of some of the major stakeholder organizations and companies to participate in strenuous negotiations, which lasted from the beginning of August through early October. These negotiations were conducted under the leadership of Senator Hatch's Counsel for Intellectual Property, Edward Damich, and with the

139. See Memorandum from William Michael Treanor, U.S. Dept. of Justice, Office of the Deputy Ass't. Att'y. Gen., to William R. Marshall, Associate White House Counsel, "Constitutional Concerns Raised by the Collections of Information Antipiracy H.R. 2652" (July 28, 1998) (on file with authors) [hereinafter DOJ Memorandum]. See also Letter from Professor Marci Hamilton to Chairman Howard Coble (Feb. 10, 1998) (on file with authors) (detailing serious Constitutional concerns) [hereinafter Letter from Hamilton].

140. See Letter from Robert Pitofsky, Chair of the Federal Trade Commission, to Tom Bliley, Chair of the House Committee on Commerce (Sept. 28, 1998) (on file with authors).

141. *Id.* at 7, 14.

142. See U.S. PATENT AND TRADEMARK OFFICE, REPORT ON (AND RECOMMENDATIONS FROM) APRIL 1998 CONFERENCE ON DATABASE PROTECTION AND ACCESS ISSUES (1998).

participation of the counterpart staffer in Senator Leahy's office, Marla Grossman.

1. Clarifying the Demands on Scientific and Technical Users

The U.S. Academies took the unusual step of participating directly in these negotiations.¹⁴³ They submitted a series of alternative proposals aimed at providing a balanced piece of legislation that would protect publishers against free-riding conduct and preserve the incentive to invest (through a true unfair competition approach) without creating a strong exclusive property right in collections of data and factual information.¹⁴⁴

Although the direct negotiations produced no major breakthroughs or compromise solutions, they did succeed in clarifying the different positions. It seems fair to say that, when exposed to direct interrogation, the publishers' detailed demands more than justified the scientific and educational communities' initial concerns.¹⁴⁵ Indeed, in response to one hypothetical situation after another, the publishers' representatives made it clear that the exclusive property right they championed for the digital network system would in fact engender the kind of legal and contractual demands on scientific, technical, and commercial users of protected databases that critics of the proposed legislation had been fearing and that are described in this article.

143. The authors of this article represented the National Academy of Sciences (NAS), the National Academy of Engineering (NAE), and the Institute of Medicine (IOM) during these negotiations.

144. See, e.g., National Academy of Sciences et al., Proposed Amendments to H.R. 2281: Synopsis, Corrections, and Text, submitted for consideration by the Senate Committee on the Judiciary (Aug. 11, 1998) (unpublished manuscript, on file with authors) [hereinafter NAS, Synopsis]; National Academy of Sciences et al., Proposed Amendments to H.R. 2281: Explanatory Memorandum (Part I) (Aug. 13, 1998) (unpublished manuscript, on file with authors) [hereinafter NAS, Explanatory Memorandum]; NAS et al., Opponents' Revised Amendments to H.R. 2281 (Sept. 4, 1998) (unpublished manuscript, on file with authors) (concerning (1) permitted acts for scientific, educational, and research purposes; (2) exclusions; (3) definition of "collections of information"; and (4) licensing).

145. For the latest iteration of the publishers' position, see *Hearings, supra* note 57 (testimony of Daniel C. Duncan, Vice President for Gov't. Affairs, Software and Information Industry Assoc.) [hereinafter Testimony of Duncan]; *Hearings, supra* note 57 (testimony of Marilyn Winokur, Exec. Vice President, Micromedex for Coalition Against Database Piracy); *Hearings, supra* note 57 (testimony of the National Assoc. of Realtors). See also *Hearings, supra* note 57 (statement of Michael K. Kink, AIPLA); *Hearings, supra* note 57 (testimony of the Agricultural Publishers Assoc.).

Perhaps because the publishers' actual demands amply confirmed the concerns that the Administration's own position papers had expressed, the final phases of the negotiations, as mediated by the Senate staffers, produced far-reaching modifications to the database component of H.R. 2281. On January 19, 1999, Senator Orrin Hatch placed in the Congressional Record a statement acknowledging that "considerable progress" had been made during the aforementioned negotiations and that, "in the end we were close to a workable compromise."¹⁴⁶ Senator Hatch also put in the Record "a discussion draft that is identical to the last of the discussion drafts ... [he had] offered last year."¹⁴⁷

2. *Compromise Proposals*

The changes incorporated in the last Discussion Draft substantially reflected the Academies' own position. Although there is no way of knowing the degree of assent to all the various provisions it contained, it is worth reviewing the package of compromise proposals embodied in the last Hatch Database Discussion Draft of October 5, 1998.¹⁴⁸

First, the strong property right approach was nudged closer to a true "misappropriation" (unfair competition) approach. This was accomplished by conditioning liability on acts that "cause *substantial* harm to the actual or neighboring market" of database proprietors,¹⁴⁹ and by inviting courts, in the draft legislative history, to determine "substantial harm" in light of "whether the harm is such as to significantly diminish the incentive to invest in gathering, organizing, or maintaining the database."¹⁵⁰

Second, a full exception that would immunize customary scientific activities was adopted by the Senate staffers, in place of the limited and unacceptable "fair use" approach that the Administration had eventually rec-

146. See 106 CONG. REC. S.316 (daily ed. Jan. 19, 1999) (statement of Senator Hatch on Database Antipiracy Legislation) [hereinafter Statement of Hatch].

147. See *id.* at S.322-26 (Chapter 14—Protection of Databases) [hereinafter Hatch Database Discussion Draft].

148. See *id.* An abridged selection of these points was used in a letter from the Presidents of the NAS, NAE, and IOM to Howard Coble, Chair of the Subcommittee on Courts and Intellectual Property of the House Committee on the Judiciary, March 9, 1999, and subsequently in the testimony of Joshua Lederberg before the Subcommittee at its March 18, 1999 Hearing. See *supra* notes 57, 114

149. See Hatch Database Discussion Draft, *supra* note 147, § 1302.

150. Hatch Database Discussion Draft Proposed, Conference Report Language, § 1302, at 33 (on file with authors).

ommended.¹⁵¹ A “fair use” approach, modeled on copyright law, would fail because other basic copyright immunities and exceptions, especially the idea-expression dichotomy, would not carry over into the database protection environment. On the contrary, because a database law protects collections of facts and data that are ineligible under copyright laws (and because scientists perceive no valid distinction between “data” and a “collection of data” in a dynamic electronic database), basic research methods that were previously permissible would become infringing acts under such a law. The burden would then fall on scientists and engineers to show that a vague fair use exception should excuse some of these infringing acts from whatever test of harm was adopted.

In contrast, the Academies successfully argued that customary and traditional scientific activities should remain untouched and unhampered by any new database protection law, exactly as the government’s initial position paper had maintained.¹⁵² To this end, section 1304 of the final version of the Hatch Database Discussion Draft stated the following proposition:

Nothing in this chapter shall prohibit or otherwise restrict the extraction or use of a database protected under this chapter for the following purposes:

- 1) for illustration, explanation, or example, comment or criticism, internal verification, or scientific or statistical analysis of the portion used or extracted; and
- 2) in the case of nonprofit scientific, educational, or research activities by nonprofit organizations, for similar *customary or traditional purposes*.¹⁵³

Only if scientists, engineers or educators working at nonprofit organizations caused substantial harm to the database-maker by using unreasonable and non-customary amounts of the collection for a given purpose, or if they in fact produced a market substitute for the original, or otherwise sought to avoid paying for the use of research tools devised as such, would

151. See Administration’s Letter 1998, *supra* note 114; see also Memorandum from the Administration to the Senate Committee on the Judiciary, *Senate Draft Developments and Suggested Additional changes to Address Key Concerns for Discussion* (rev. Sept. 30, 1998) (on file with authors).

152. See Administration’s Letter 1998, *supra* note 114. However, the Government’s latest position paper leans towards the compromise “fair use” provision introduced in H.R. 354, which we judge to be inadequate. See Statement of Pincus, *supra* note 87, at 29-31.

153. Hatch Database Discussion Draft, *supra* note 147, § 1304(a) (emphasis added).

liability kick in.¹⁵⁴ On this approach, the burden fell on publishers to show that scientists had crossed the line of permitted, traditional, or customary uses, which were otherwise immunized. The guiding principle that science, technology, and education should be left no worse off after enactment than they were before, as proposed by ICSU,¹⁵⁵ would thus have been implemented.

Third, additional immunities and exceptions favoring certain instructional and library uses of databases were also defined,¹⁵⁶ although more thought needs to be given to educational users generally in this context.¹⁵⁷ Fourth, efforts were also made to reduce the likelihood that private interests might permanently capture government-generated data,¹⁵⁸ although more remains to be done on this score as well.¹⁵⁹

Fifth, a clearly-worded duration clause ending protection after fifteen years reduced (but did not altogether eliminate) the risk of perpetual protection.¹⁶⁰ A rudimentary database deposit scheme was also proposed, which increased the likelihood of data eventually entering the public do-

154. *See id.* § 1304(b):

In no case may a use or extraction for a purpose described in subsection (a) be permitted if the substantial harm referred to in Section 1302—

1. arises because the amount of the portion used or extracted is more than is reasonable and customary for the purpose;

2. consists of the use or extraction being intended to, or being likely to serve as a substitute for or to supplant all or a substantial part of the database from which the extraction or use is made or an adaptation thereof that is protected under this chapter;

3. arises because the extraction or use is intended to avoid payment of reasonable fees for use of a database incorporated into a product or service specifically marketed for educational, scientific, or research purposes; or

4. arises because the use or extraction is part of a pattern, system, or repeated practice by the same party, related parties, or parties acting in concert with respect to the same database or a series of related databases.

155. *See* ICSU Position Paper, *supra* note 34, at 2.

156. *See* Hatch Database Discussion Draft, *supra* note 147, § 1307.

157. *See also* Statement of Phelps, *supra* note 96, at 16 (proposing exemption for non-profit teaching activities).

158. *See* Hatch Database Discussion Draft, *supra* note 147, § 1305(a)-(c).

159. *See* Statement of Pincus, *supra* note 87, at 13-20 (detailed proposals concerning government-funded data and the need to avoid its "capture" by commercial interests); Ginsburg, *supra* note 85, at 27-29.

160. *See* Statement of Pincus, *supra* note 87, at 24-27.

main, albeit in a cumbersome and, perhaps, costly fashion.¹⁶¹ If that route were taken, more incentives would be needed to ensure that deposits were actually made. However, the Administration favors developing other, simpler incentives to ensure the availability of public domain data that are worth exploring.¹⁶²

Sixth, the need for some regulation of licensing terms and conditions was expressly recognized. A series of provisions required periodic studies of the misuse doctrine as applied to licensing agreements, or to the use of technological measures that might frustrate the "permitted acts" clause of the bill. Particular grounds for the study included sole-source provider contracts that imposed unreasonable terms or conditions; tying or other practices traditionally recognized as abusive; and practices shown to have "prevented access to valuable information for research, competition, or innovation purposes."¹⁶³ The draft legislative history then clarified that courts were free to apply these same criteria to claims of misuse arising after the time of enactment and need not "refrain from applying the doctrine of misuse until the study is completed."¹⁶⁴ There was some further possibility that criteria for evaluating the misuse of licensing agreements might have ultimately been codified in the operative clauses of the Act itself.

Finally, the draft legislative history to accompany these measures also clarified the definition of databases in ways that tended to exclude ordinary literary works, and it denied protection "to any ideas, facts, procedure, process, system, method of operation, concept, principle, or discovery, as distinct from the collection that is the product of investment protected by this Act."¹⁶⁵ Needless to say, we think the proposed database law should expressly codify these provisions. Indeed, the fact that the bill's proponents oppose inclusion of such a basic limitation in H.R. 2281 only serves to reinforce our concerns about the true nature and extent of their intended exploitation of the legislation's most restrictive provisions.

161. See Hatch Database Discussion Draft, *supra* note 147, § 1310(d); Statement of Pincus, *supra* note 87, at 24-27.

162. See Statement of Pincus, *supra* note 87, at 24-27 (discussing dated identity tags and statutory defenses to liability, including lack of public availability of government-funded data); see also Ginsburg, *supra* note 85, at 24.

163. Hatch Database Discussion Draft, *supra* note 147, § 4. See also *id.* § 1306(a) (preserving doctrine of misuse). But see *id.* § 1306(e) (allowing licensors unrestricted freedom of contract).

164. Hatch Database Discussion Draft, Proposed Conference Report Language, *supra* note 150, at 36-37.

165. *Id.* at 131.

C. Uncertain Future of the Database Protection Law

The foregoing discussion reveals the extent to which the Hatch Database Discussion Draft evolved away from the strong exclusive property right approach, adopted in the E.U. Directive, toward a more balanced unfair competition approach that protected publishers against piracy while consciously avoiding harm to science, education, and other public-good uses of data. Of course, not all the issues of concern to science were addressed in a fully satisfactory manner; but given the need for compromise and consensus, the ability of the staff to produce a relatively balanced bill from such unpromising material as the House bill deserves commendation.

Perhaps the biggest unaddressed issue was that of value-adding uses. The Discussion Draft did not resolve the tensions between a dominant group of database publishers, who seek to control value-adding uses of protected collections, and a dissident group of publishers and allies, who believe value-adding uses should remain as unfettered as possible.¹⁶⁶ On this point, the Academies proposed a scheme favoring easy use of data for commercial value-adding purposes in exchange for the payment of reasonable royalties under an automatic licensing scheme,¹⁶⁷ but neither side would accept this approach. Nevertheless, under the misappropriation approach to "substantial harm," as elaborated in the Draft Legislative History,¹⁶⁸ courts could work out the criteria for balancing incentives to invest against incentives to compete for the short run, and these case-by-case solutions could be legislatively evaluated later on.

In the end, the Hatch Discussion Draft was not adopted mainly because time ran out in which to remove the last remaining wrinkles that prevented an agreed compromise.¹⁶⁹ As a result, the database component of H.R. 2281, Title V, was stripped from the Digital Millennium Copyright Act, which was enacted at the end of the legislative year. Work on

166. Compare, e.g., Testimony of Duncan, *supra* note 145, at 9-10 (presenting a case for broad scope of protection) with Testimony of CCIA et al., *supra* note 101, at 6-8 (advocating less restricted commercial reuse of data).

167. See NAS, Synopsis, *supra* note 144, § 1303(g) (distinguishing use by competitors on distant markets from use in direct competition for these purposes). See also Reichman & Samuelson, *supra* note 7, at 146 (proposing "automatic license built into ... [a] modified liability right itself"); Ginsburg, *supra* note 85, at 28 (proposing automatic license for sole-source collections of government data to benefit "market rivals").

168. See *supra* note 150.

169. See *supra* note 146 and accompanying text (expressing Senator Hatch's belief that considerable progress had been made).

database protection has begun all over again under the aegis of a new Congress, which convened in January 1999.¹⁷⁰

During the fall of 1998, some members of the coalition that had opposed H.R. 2281 drafted still another bill that sought to implement unfair competition principles more aggressively than was contemplated in the final Hatch Discussion Draft.¹⁷¹ There exists some support for this so-called "minimalist" unfair competition approach, which would protect databases only against wholesale duplication for an indefinite period of time.¹⁷² However, this solution could easily degenerate into a *de facto* exclusive property right conferring perpetual protection by the back door.

Whatever happens next, the final version of the Hatch Discussion Draft constituted a milestone along the route towards a more balanced model of database protection, and its lessons should inform the next round of legislative deliberations. There is unofficial and anecdotal evidence that the Japanese government may also embark upon a true unfair competition approach, which, if true, would afford a unique opportunity for the United States and Japan to present a united front to the rest of the world. In that event, other countries would probably move in the direction of a more balanced unfair competition regime, which might leave the E.U. alone to continue its experiment with a strong property right, or to modify its Directive so as to obtain a more balanced system of protection with fewer social costs.

Unfortunately, the scientific community will experience serious challenges to the policy of easy access to, and unrestricted uses of data, regardless of the approach to database protection that ultimately emerges from Congress and the legislation of other countries. As pointed out at the beginning of this article, publishers can already control the dissemination of data by combining technical protection measures with adhesion contracts in the online environment, even without the adoption of specific database legislation.¹⁷³ While the presence of an intellectual property right would strengthen the publishers' position and put the scientific and technical communities under grave legal disadvantages,¹⁷⁴ the absence of an ex-

170. See H.R. 354, 106th Cong. (1999); *supra* text accompanying notes 57-61.

171. See Statement of Hatch, *supra* note 146, at S.320-22 (Proposed Bill to Amend Title 17, United States Code, to Promote Research and Fair Competition in the Database Industry) [hereinafter Minimalist Protection Bill of 1999].

172. See Minimalist Protection Bill of 1999, *supra* note 171, §§ 1401, 1408.

173. See *supra* text accompanying notes 10-18.

174. See Reichman & Franklin, *supra* note 9, at 911-14, 947-51 (exploring impact of online adhesion contracts with or without a codified property right underneath).

clusive property right would not free them from the need to rethink their whole approach to maintaining the unrestricted flow of scientific and technical data in an emerging information economy.

We trust that two new studies by the National Research Council, which are examining these issues, will shed further light on the options for science when they are published in mid-1999. Meanwhile, it seems clear that the scientific and technical communities will have to consider ways of reconciling a greater degree of commercialization for databases generated within the academic community with the need to maintain privileged access to the same databases for scientific and other public interest objectives. Universities and research institutions that generate data will thus have to develop rules for disciplining grants and the uses of data obtained from grants. Separate channels for the nonprofit distribution of scientific and technical data may have to be created, with particular rules for participating organizations. Ultimately, it may also prove desirable to develop an extended licensing authority for certain classes of scientific data, in order to administer these resources with low transaction costs and uniform rules for commercial and non-commercial users.¹⁷⁵

In general, efforts must be made to preserve the sharing ethos with respect to publicly-generated scientific data, to encourage those who invest in the production of privately-generated data to provide price discrimination in favor of the scientific and educational communities, and to develop differentiated products for the non-profit sector.¹⁷⁶ As the Academies recently explained, Congress should strive to reconcile legitimate measures to repress parasitical copying of protected databases with the equally legitimate needs of the scientific, technical, and educational communities. These communities require:

- access to data on fair and reasonable conditions;
- the ability to use the data accessed for research or educational purposes; and
- freedom from contractual or technical interference with these pursuits.¹⁷⁷

These objectives will, in turn, require close collaboration with governments. The goal is to ensure that data generated at the taxpayers' ex-

175. *Cf. Merges, supra* note 1.

176. *See BITS OF POWER, supra* note 69, at 166-71.

177. *See NAS, Explanatory Memorandum, supra* note 144, at 9.

pense remain available at least for scientific and educational purposes, and that efforts to stimulate greater investment in the development of new databases do not end by creating barriers to entry or otherwise discouraging follow-on innovation and public good uses of the building blocks of knowledge.

IV. PRESERVING THE CONSTITUTIONAL BALANCE OF INTERESTS IN THE NETWORKED ENVIRONMENT

Because everything on the Internet is potentially a “database” or a “collection of information” in our increasingly information-based economy, the law that protects collected information will determine the level of competition and prices in that economy. The EU Directive—and to almost the same extent its counterpart proposal pending in the United States¹⁷⁸—opt for a very high level of protection. These regimes buttress mini-monopolies of data and information that could threaten the advance of scientific and technical research, hinder the creation of legitimate new commercial information products, and hurt downstream consumer interests.

A. The Competitive Ethos Under Attack

The fallacy behind most proposals for strong forms of database protection is that they ignore the dual nature of data and information as such. On one level, data function as a raw material of the information economy, a basic ingredient of the public domain, from which scientists and entrepreneurs both draw to fashion their respective products. On a second level, data and information are bundled into downstream products that attract intellectual property rights and related contractual agreements. The mistake is to presume that strong intellectual property rights that were empirically well-suited to downstream applications—mainly derived from the patent and copyright models—are equally well-suited to upstream regulation of the data as inputs into the process of innovation.

The opposite is true. If we balkanize the public domain and make the transaction costs of recreating it by contracts prohibitively expensive and complex, a dysfunctional legal system will impede the cumulative and sequential development of technical paradigms by depriving routine innovators of access to the building blocks of knowledge.¹⁷⁹

178. See E.U. Directive, *supra* note 6; H.R. 354, 106th Cong. (1999).

179. See Heller & Eisenberg, *supra* note 130; Suzanne Scotchmer, *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*, 5 J. ECON. PERSPECTIVES 29 (1991); Reichman & Franklin, *supra* note 9, at 884-88. Cf. generally Michael A. Hel-

The truth is that traditionally we have left small grain-size innovation to weaker forms of entitlement, that is, to liability principles rooted in unfair competition law, rather than strong property rights, and this has been a basic premise on which the competitive economy of the industrial revolution was constructed.¹⁸⁰ These lessons are still germane to the information economy—lessons sounding in reverse engineering and the reuse of ideas, rather than in legally supported monopolies on products of routine innovation and investment.

Until convincing evidence to the contrary accrues, we should address the risk of market failure in the information economy by erring on the side of underprotection rather than overprotection. This follows because there is no real or potential shortage of investment in this milieu once the causes of market failure are controlled; and it is sound public policy, because we do not wish needlessly to encourage the monopolization of the sources of factual data, to deter value-adding innovators, or to retard the progress of science.¹⁸¹

B. The Constitutional Dilemma

The inclination to place strong intellectual property rights in upstream collections of information is contrary to our entire intellectual property tradition and to our basic constitutional heritage. For some forty years, the late Professor Melville Nimmer, a leading authority on both copyright and First Amendment law, taught that copyright protection would violate First Amendment guarantees of free speech were it not for the judicial exclusion of ideas and facts from the reach of the exclusive property rights granted to authors and artists.¹⁸² In 1976, Congress codified that exclusion

ler, *The Tragedy of the Anticommons: Property in the Transition from Marx to Markets*, 111 HARV. L. REV. 621 (1998) (showing that too many rights to exclude produce anti-commons effects as bad as the lack of any powers to exclude).

180. See generally Reichman, *Legal Hybrids*, *supra* note 3, at 2434-44; Reichman, *Charting*, *supra* note 3, at 485-96, 504-17.

181. See generally J.H. Reichman, *Solving the Green Tulip Problem: Repackaging Rights in Subpatentable Innovation*, paper presented to New York University Conference on "Intellectual Products: Novel Claims to Protection and Their Boundaries," Engelberg Center on Innovation Law and Policy, La Pietra, Italy (June 25-28, 1998) (unpublished manuscript, on file with authors).

182. See M.B. NIMMER & D. NIMMER, *NIMMER ON COPYRIGHT* § 1.10 (1998); R.A. SMOLLA, *SMOLLA & NIMMER ON FREEDOM OF SPEECH* § 21:8 (1998); Robert C. Denicola, *Copyright and Free Speech: Constitutional Limitations on the Protection of Expression*, 67 CALIF. L. REV. 283 (1983); M.B. Nimmer, *Does Copyright Abridge the First Amendment Guarantees of Free Speech and Press?*, 17 UCLA L. REV. 1180 (1970); Al-

in Section 102(b) of the General Revision of Copyright Law,¹⁸³ and in 1991, the Supreme Court, in *Feist Publications, Inc. v. Rural Telephone Service, Co.*,¹⁸⁴ reconfirmed the constitutional prohibition against an exclusive property right in either facts or ideas.

Proponents of H.R. 354 and its predecessors openly concede that the "harm to actual or potential markets" test was drawn from Section 107(4) of the 1976 Copyright Law, which codified its fair use provisions.¹⁸⁵ This is a constitutionally dubious admission because the very purpose of Section 107(4) is to confirm that protection of the author's market interests in both primary and secondary markets constitutes the true goal of the copyright law's exclusive rights, exactly as Judge Frank declared in his famous opinion in *Arnstein v. Porter*.¹⁸⁶

When transplanted to the database milieu, however, the protection of mere investment in databases that do not rise to the level of creative works of authorship against "harm to actual or potential markets"¹⁸⁷ indirectly creates an exclusive property right in noncopyrightable collections of data, which governs both primary and secondary markets. Once collected, no one can make further use of the facts and data contained in the collection without the compiler's permission, even though Section 102(b) of the Copyright Law states that facts and ideas are not fit subjects of an exclusive property right.

True, H.R. 354 does allow "independent creation" of databases in Section 1403(c) and Section 1403(a) exempts nonprofit educational, scientific, and research uses from liability for causing harm to "potential" markets and for certain other reasonable uses.¹⁸⁸ But such defenses in the proposed database regime are no more curative of these constitutional flaws than they would be in the copyright regime, *for the reason that no one can constitutionally oblige all persons not to use facts or ideas that*

fred C. Yen, *A First Amendment Perspective on the Idea/Expression Dichotomy and Copyright in a Work's "Total Concept and Feel,"* 38 EMORY L.J. 393 (1989).

183. See 17 U.S.C. § 102(b) (1994).

184. 499 U.S. 340 (1991).

185. 17 U.S.C. § 107(4) (1994). See also *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984); PATRY, *supra* note 30, at 205-10.

186. 154 F.2d 464 (2d. Cir. 1946), *cert. denied*, 330 U.S. 851 (1947). See generally PAUL GOLDSTEIN, COPYRIGHT §§ 7.1.1.2, 7.4.1.2 (1998) (discussing infringement in terms of protection of authors' market interest).

187. H.R. 354, 106th Cong. § 1402 (1999).

188. See *id.* §§ 1403(c), 1403(a)(1) (1999) (listing "certain nonprofit educational, scientific, or research uses"); see also *id.* § 1403(a)(2) (listing "Additional Reasonable Uses").

have been made available to the public. Facts and ideas that the copyright law must leave to unrestricted public use cannot constitutionally be withdrawn from public use under the First Amendment by a database law that protects against extraction and use on both primary and derivative markets.¹⁸⁹

In this connection, one should recall that the copyright law, unlike the patent law, does not protect against *use* as such of even the protected expression, as the Supreme Court established in *Baker v. Selden*.¹⁹⁰ The protection of noncopyrightable data and facts against *use* on both primary and secondary markets thus impermissibly disrupts the balance established in the federal copyright and patent laws, which implement the constitutional Enabling Clause.¹⁹¹ Notwithstanding the public's right to use facts and ideas under the First Amendment and notwithstanding the constraints limiting Congressional action under the constitutional Enabling Clause, H.R. 354 institutes copyright-like protection for the use of noncopyrightable matter, creates a *de facto* derivative work right in noncopyrightable compilations, and prohibits transformative uses—however pro-competitive in nature—that harm this reserved or derivative market on the “potential harm” test.

No invocation of unfair competition law can disguise the fact that a “harm to actual or potential markets” test that does not focus on unfair or improper conduct expresses the language of exclusive property rights, which is exactly the function that Section 107(4) performs in the Copyright Law.¹⁹² The fact that it is often physically or economically impracticable to regenerate scientific and research data from scratch only enhances the potential restraints on free speech under H.R. 354 as it stands, by risk-

189. See Letter from Hamilton, *supra* note 139. See also Malla Pollack, *The Right to Know? Delimiting Database Protection at the Juncture of the Commerce Clause, the Intellectual Property Clause, and the First Amendment*, 17 CARDOZO ARTS & ENT. L. J. 47, 67-74 (1999).

190. See *Baker v. Selden*, 101 U.S. 99 (1879); Brown, *supra* note 28. Only those uses specified in the Act are protected. 17 U.S.C. §§ 106, 106A (1994).

191. See U.S. CONST. art. I, § 8, cl. 8; 17 U.S.C. § 301 (1994) (preemption); *Bonito Boats Inc. v. Thunder Craft Boats Inc.*, 489 U.S. 141 (1989); *Compco Corp. v. Day-Brite Lighting Inc.*, 376 U.S. 234 (1964); *Sears, Roebuck & Co. v. Stiffel Co.*, 376 U.S. 225 (1964).

192. 17 U.S.C. § 107(4) (1994); *supra* note 185; see also Letter from Professor Harvey Perlman, Co-Reporter, RESTATEMENT (THIRD) OF UNFAIR COMPETITION LAW (1994), to Senator Orrin Hatch, Chairman of the Senate Committee on the Judiciary (Sept. 8, 1998) (on file with authors) (explaining why neither H.R. 354 nor its predecessors are, in fact, unfair competition laws) [hereinafter Perlman letter].

ing the withdrawal of facts and data as such from the public domain.¹⁹³ The federal appellate courts have consistently declared that avoiding the costs of regenerating known facts and ideas constitutes a basic economic premise underlying the constraints on intellectual property protection deriving from both the First Amendment and the Enabling Clause.¹⁹⁴

The broad definitions of both "collection of information"¹⁹⁵ and "information"¹⁹⁶ in H.R. 354 aggravate these constitutional infirmities by drawing "works of authorship" into the realm of a competing and overlapping intellectual property right, and also by casting legal doubts upon the future ability of third parties to make untrammelled use of public domain matter. Anytime someone would use data, including historical data, that are made available to the public contained in a "collection of information" protected by the proposed law, that user would be exposed to claims that he or she will have harmed the database originator's actual or potential markets if that producer had also used the same or similar data. This broad risk of liability cannot fail to have a chilling effect on the use of known facts and noncopyrightable databases in both the commercial and non-commercial spheres; and it is of little consolation to researchers and educators that they must fear only harm they might cause to actual markets, rather than to potential markets, as well.¹⁹⁷

C. Erring on the Side of Caution

In contrast, a true unfair competition approach would attach liability only when the third party harmed the database maker's actual or potential

193. See *supra* text accompanying notes 120-21.

194. See, e.g., *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991); *Harper & Row v. Nation Enter.*, 471 U.S. 539 (1985). As the Department of Justice's Office of the Legal Counsel recently affirmed, to the extent that the proposed legislation would prohibit extractions or uses of substantial portions of factual compilations by direct competitors, it is much more likely to be held constitutional than if it would prohibit extractions or uses by potential consumers for noncommercial purposes. By contrast, if the provision were construed to provide protection against uses by potential consumers, and not simply direct competitors, it would appear to be of almost limitless scope and therefore to raise constitutional concerns that would appear insurmountable.

DOJ Memorandum, *supra* note 139, at 8.

195. See H.R. 354, 106th Cong. §1401(1) (1999).

196. See *id.*, §1401(2).

197. See *id.*, §1403(a)(1).

market *by improper, unfair, or dishonest means*.¹⁹⁸ Such an approach would not inhibit competitors who “harm” the market by honest and innovative means, and it would not impede true transformative uses that promote competition and the public interest in science and education.¹⁹⁹

The “actual or potential markets” test is thus so broad that it would hinder fair competition simply because every successful competitor harms a prior entrant’s market by definition and because would-be competitors would never know in advance when the use or extraction of protected data may turn out to cause harm to some unknown potential market. In this and other respects, the “harm to markets” test actually cloaks a reserved market formula, in the manner of the exclusive rights to reproduce and to prepare derivative works granted by the Copyright Law.²⁰⁰ Use of this formula in the database context invites other industries to apply for similar protection against harm to their actual or potential markets; and the cumulative anti-competitive effects of recognizing such special-interest protectionist pleas could seriously undermine the ability of the United States to compete in an integrated global marketplace.

United States intellectual property law and policy have traditionally mandated the unfettered use of noncopyrightable facts and of subpatentable ideas, and have favored unbridled competition with respect to the products of mere investment.²⁰¹ We shall undoubtedly experience suboptimal investment in the production of databases if Congress fails to protect publishers against certain forms of piratical conduct that threaten to deprive them of the fruits of their investment.²⁰² But if we combat this risk of market failure by enforcing strong monopolies in collections of data, we may end up balkanizing the information economy by recreating the medieval economic quandary in which products could not flow across countries or continents because too many feudal monopolists demanded payments every few miles down the road.²⁰³

If we discourage follow-on innovation and public good uses of the very databases whose development statutory legal protection is supposed

198. Compare RESTATEMENT (THIRD) OF UNFAIR COMPETITION LAW § 38 (1996) (rejecting general misappropriation doctrine) and Perlman letter, *supra* note 192 with Gordon, *supra* note 7 (proposing tort of malcompetitive copying).

199. Cf. Reichman & Samuelson, *supra* note 7, at 137-45.

200. See 17 U.S.C. §§ 106(1)-(2) (1994).

201. See Reichman, *Charting*, *supra* note 3, at 485-96 (“Negative Economic Premises Underlying the Dominant Legal Paradigms”); *supra* note 191.

202. See, e.g., Tyson & Sherry, *supra* note 72; Hunsucker, *supra* note 7.

203. Cf. Heller & Eisenberg, *supra* note 130; Heller, *supra* note 179.

to stimulate, the end result may be bad for the database industry as a whole²⁰⁴ and devastating for our whole scientific and technical innovation system, which depends on the relatively unrestricted flow of factual data. Instead, we need a regime that loosely preserves a balanced relationship between public and private interests, which courts can develop gradually in response to the empirical conditions of the evolving information economy.²⁰⁵

In the Information Age, as in the Industrial Revolution, we should continue to believe that competition is the lifeblood of commerce, and we should accordingly structure all legal entitlements so as to produce a high degree of competition and maximum dissemination of data and information. If we err on the side of caution and underprotect the building blocks of knowledge, we can always adjust the level of protection upwards later on, in the face of compelling empirical evidence of real economic harm.

But the opposite is not true. Acquired rights and legislatively enacted monopolies cannot easily be eradicated. The wrong decisions today could lessen the vitality of our research enterprise, weaken the national system of innovation, and compromise our future technological superiority, which all depend on maintaining an appropriate balance between upstream and downstream uses of data and factual information.

204. See Testimony of CCIA et al., *supra* note 101.

205. See also Pamela Samuelson (unpublished and untitled essay, manuscript on file with authors) (explaining the need for a national information policy detached from intellectual property policies); Litman, *supra* note 16; Reichman & Franklin, *supra* note 9, at 968-70.