AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

# DATABASE AND COLLECTIONS OF INFORMATION MISAPPROPRIATIONS

## JOINT HEARING

BEFORE THE

SUBCOMMITTEE ON COURTS, THE INTERNET, AND INTELLECTUAL PROPERTY

OF THE

COMMITTEE ON THE JUDICIARY

AND THE

SUBCOMMITTEE ON COMMERCE, TRADE, AND CONSUMER PROTECTION

OF THE

COMMITTEE ON ENERGY AND COMMERCE

HOUSE OF REPRESENTATIVES

ONE HUNDRED EIGHTH CONGRESS

FIRST SESSION

SEPTEMBER 23, 2003

**Serial No. 51**
(Committee on the Judiciary)

**Serial No. 108–46**
(Committee on Energy and Commerce)

Printed for the use of the Committee on the Judiciary
and the Committee on Energy and Commerce

Available via the World Wide Web: http://www.house.gov/judiciary
and http://energycommerce.house.gov

U.S. GOVERNMENT PRINTING OFFICE

89–463 PDF WASHINGTON : 2003

For sale by the Superintendent of Documents, U.S. Government Printing Office
Internet: bookstore.gpo.gov Phone: toll free (866) 512–1800; DC area (202) 512–1800
Fax: (202) 512–2250 Mail: Stop SSOP, Washington, DC 20402–0001

## COMMITTEE ON THE JUDICIARY

F. JAMES SENSENBRENNER, JR., Wisconsin, *Chairman*

HENRY J. HYDE, Illinois
HOWARD COBLE, North Carolina
LAMAR SMITH, Texas
ELTON GALLEGLY, California
BOB GOODLATTE, Virginia
STEVE CHABOT, Ohio
WILLIAM L. JENKINS, Tennessee
CHRIS CANNON, Utah
SPENCER BACHUS, Alabama
JOHN N. HOSTETTLER, Indiana
MARK GREEN, Wisconsin
RIC KELLER, Florida
MELISSA A. HART, Pennsylvania
JEFF FLAKE, Arizona
MIKE PENCE, Indiana
J. RANDY FORBES, Virginia
STEVE KING, Iowa
JOHN R. CARTER, Texas
TOM FEENEY, Florida
MARSHA BLACKBURN, Tennessee

JOHN CONYERS, JR., Michigan
HOWARD L. BERMAN, California
RICK BOUCHER, Virginia
JERROLD NADLER, New York
ROBERT C. SCOTT, Virginia
MELVIN L. WATT, North Carolina
ZOE LOFGREN, California
SHEILA JACKSON LEE, Texas
MAXINE WATERS, California
MARTIN T. MEEHAN, Massachusetts
WILLIAM D. DELAHUNT, Massachusetts
ROBERT WEXLER, Florida
TAMMY BALDWIN, Wisconsin
ANTHONY D. WEINER, New York
ADAM B. SCHIFF, California
LINDA T. SÁNCHEZ, California

PHILIP G. KIKO, *Chief of Staff-General Counsel*
PERRY H. APELBAUM, *Minority Chief Counsel*

---

## SUBCOMMITTEE ON COURTS, THE INTERNET, AND INTELLECTUAL PROPERTY

LAMAR SMITH, Texas, *Chairman*

HENRY J. HYDE, Illinois
ELTON GALLEGLY, California
BOB GOODLATTE, Virginia
WILLIAM L. JENKINS, Tennessee
SPENCER BACHUS, Alabama
MARK GREEN, Wisconsin
RIC KELLER, Florida
MELISSA A. HART, Pennsylvania
MIKE PENCE, Indiana
J. RANDY FORBES, Virginia
JOHN R. CARTER, Texas

HOWARD L. BERMAN, California
JOHN CONYERS, JR., Michigan
RICK BOUCHER, Virginia
ZOE LOFGREN, California
MAXINE WATERS, California
MARTIN T. MEEHAN, Massachusetts
WILLIAM D. DELAHUNT, Massachusetts
ROBERT WEXLER, Florida
TAMMY BALDWIN, Wisconsin
ANTHONY D. WEINER, New York

BLAINE MERRITT, *Chief Counsel*
DEBRA ROSE, *Counsel*
DAVID WHITNEY, *Counsel*
MELISSA L. MCDONALD, *Full Committee Counsel*
ALEC FRENCH, *Minority Counsel*

## COMMITTEE ON ENERGY AND COMMERCE

W.J. "BILLY" TAUZIN, Louisiana, *Chairman*

MICHAEL BILIRAKIS, Florida
JOE BARTON, Texas
FRED UPTON, Michigan
CLIFF STEARNS, Florida
PAUL E. GILLMOR, Ohio
JAMES C. GREENWOOD, Pennsylvania
CHRISTOPHER COX, California
NATHAN DEAL, Georgia
RICHARD BURR, North Carolina
*Vice Chairman*
ED WHITFIELD, Kentucky
CHARLIE NORWOOD, Georgia
BARBARA CUBIN, Wyoming
JOHN SHIMKUS, Illinois
HEATHER WILSON, New Mexico
JOHN B. SHADEGG, Arizona
CHARLES W. "CHIP" PICKERING, Mississippi
VITO FOSSELLA, New York
ROY BLUNT, Missouri
STEVE BUYER, Indiana
GEORGE RADANOVICH, California
CHARLES F. BASS, New Hampshire
JOSEPH R. PITTS, Pennsylvania
MARY BONO, California
GREG WALDEN, Oregon
LEE TERRY, Nebraska
ERNIE FLETCHER, Kentucky
MIKE FERGUSON, New Jersey
MIKE ROGERS, Michigan
DARRELL E. ISSA, California
C.L. "BUTCH" OTTER, Idaho

JOHN D. DINGELL, Michigan
*Ranking Member*
HENRY A. WAXMAN, California
EDWARD J. MARKEY, Massachusetts
RALPH M. HALL, Texas
RICK BOUCHER, Virginia
EDOLPHUS TOWNS, New York
FRANK PALLONE, Jr., New Jersey
SHERROD BROWN, Ohio
BART GORDON, Tennessee
PETER DEUTSCH, Florida
BOBBY L. RUSH, Illinois
ANNA G. ESHOO, California
BART STUPAK, Michigan
ELIOT L. ENGEL, New York
ALBERT R. WYNN, Maryland
GENE GREEN, Texas
KAREN McCARTHY, Missouri
TED STRICKLAND, Ohio
DIANA DeGETTE, Colorado
LOIS CAPPS, California
MICHAEL F. DOYLE, Pennsylvania
CHRISTOPHER JOHN, Louisiana
TOM ALLEN, Maine
JIM DAVIS, Florida
JAN SCHAKOWSKY, Illinois
HILDA L. SOLIS, California

DAN R. BROUILLETTE, *Staff Director*
JAMES D. BARNETTE, *General Counsel*
REID P.F. STUNTZ, *Minority Staff Director and Chief Counsel*

---

## SUBCOMMITTEE ON COMMERCE, TRADE, AND CONSUMER PROTECTION

CLIFF STEARNS, Florida, *Chairman*

FRED UPTON, Michigan
BARBARA CUBIN, Wyoming
JOHN SHIMKUS, Illinois
JOHN B. SHADEGG, Arizona
*Vice Chairman*
GEORGE RADANOVICH, California
CHARLES F. BASS, New Hampshire
JOSEPH R. PITTS, Pennsylvania
MARY BONO, California
LEE TERRY, Nebraska
ERNIE FLETCHER, Kentucky
MIKE FERGUSON, New Jersey
DARRELL E. ISSA, California
C.L. "BUTCH" OTTER, Idaho
W.J. "BILLY" TAUZIN, Louisiana
(Ex Officio)

JAN SCHAKOWSKY, Illinois
*Ranking Member*
HILDA L. SOLIS, California
EDWARD J. MARKEY, Massachusetts
EDOLPHUS TOWNS, New York
SHERROD BROWN, Ohio
JIM DAVIS, Florida
PETER DEUTSCH, Florida
BART STUPAK, Michigan
GENE GREEN, Texas
KAREN McCARTHY, Missouri
TED STRICKLAND, Ohio
DIANA DeGETTE, Colorado
JOHN D. DINGELL, Michigan,
(Ex Officio)

# CONTENTS

SEPTEMBER 23, 2003

OPENING STATEMENT


Page

The Honorable Lamar Smith, a Representative in Congress From the State of Texas, and Chairman, Subcommittee on Courts, the Internet, and Intellectual Property, Committee on the Judiciary ..... 1

The Honorable Cliff Stearns, a Representative in Congress From the State of Florida, and Chairman, Subcommittee on Commerce, Trade, and Consumer Protection, Committee on Energy and Commerce ..... 2

The Honorable Howard L. Berman, a Representative in Congress From the State of California, and Ranking Member, Subcommittee on Courts, the Internet, and Intellectual Property, Committee on the Judiciary ..... 4

The Honorable Jan Schakowsky, a Representative in Congress From the State of Illinois and Ranking Member, Subcommittee on Commerce, Trade, and Consumer Protection, Committee on Energy and Commerce ..... 4


WITNESSES


Mr. David Carson, General Counsel, Copyright Office of the United States, Library of Congress

Oral Testimony ..... 9

Prepared Statement ..... 11

Mr. Thomas J. Donohue, President and Chief Executive Officer, Chamber of Commerce

Oral Testimony ..... 17

Prepared Statement ..... 18

Mr. Keith Kupferschmid, Vice President, Intellectual Property Policy and Enforcement Software and Information Industry Association, on behalf of the Coalition Against Database Piracy

Oral Testimony ..... 20

Prepared Statement ..... 23

Mr. William Wulf, President, National Academy of Engineering and Vice Chairman, National Research Council

Oral Testimony ..... 40

Prepared Statement ..... 42


LETTERS, STATEMENTS, ETC., SUBMITTED FOR THE HEARING


Letter to the Honorable F. James Sensenbrenner, Jr., Committee on the Judiciary and the Honorable W. J. "Billy" Tauzin, Committee on Energy and Commerce from Nils Hasselmo, President, Association of American Universities ..... 6

Letter from Thomas J. Donohue, President and Chief Executive Officer, Chamber of Commerce of the United States of America to the Honorable Lamar S. Smith, Chairman, Committee on the Judiciary, Subcommittee on Courts, the Internet and Intellectual Property and the Honorable Cliff Stearns, Chairman, Subcommittee on Commerce, Trade and Consumer Protection of the House Committee on Energy and Commerce ..... 66

Page

## APPENDIX

### Material Submitted for the Hearing Record

Statement of the Honorable W. J. "Billy" Tauzin ..... 79
Statement of the Honorable Bart Stupak ..... 79
Statement of the Honorable Barbara Cubin ..... 80
Statement of the Honorable Gene Green ..... 80
Statement of the Honorable Ted Strickland ..... 80
Letter from Justin Hughes, Assistant Professor of Law, Benjamin N. Cardozo School of Law, Yeshiva University ..... 81
Letter from Keith Kupferschmid, Counsel for CADP to the Honorable Lamar Smith, Chairman, Subcommittee on Courts, the Internet and Intellectual Property, Committee on the Judiciary and the Honorable Cliff Stearns, Chairman, Subcomittee on Commerce, Trade and Consumer Protection of the House Energy and Commerce Committee ..... 107
Statement of Congressman John D. Dingell, Ranking Member, Committee on Energy and Commerce ..... 151
Statement of Congressman John Shimkus ..... 152

# DATABASE AND COLLECTIONS OF INFORMATION MISAPPROPRIATIONS

**TUESDAY, SEPTEMBER 23, 2003**

HOUSE OF REPRESENTATIVES,
SUBCOMMITTEE ON COURTS, THE INTERNET,
AND INTELLECTUAL PROPERTY,
COMMITTEE ON THE JUDICIARY,

AND

SUBCOMMITTEE ON COMMERCE, TRADE,
AND CONSUMER PROTECTION,
COMMITTEE ON ENERGY AND COMMERCE,
*Washington, DC.*

The Subcommittees met, pursuant to call, at 4:08 p.m., in Room 2141, Rayburn House Office Building, Hon. Lamar Smith [Chairman of the Subcommittee on Courts, the Internet, and Intellectual Property] presiding.

Mr. SMITH. The Subcommittee on Courts, the Internet, and Intellectual Property will come to order along with the written permission of the Chairman of the Subcommittee on Commerce, Trade, and Consumer Protection.

Before I recognize individuals for opening statements, let me just make a couple of comments. To my knowledge, this is the first time we have had such a joint hearing, and it is a privilege to do so with the Commerce Committee, one, because they are so important; but, two, because Cliff Stearns, the Chairman of their Subcommittee, is a personal friend and for a number of years, actually was a neighbor across the hall. And I miss seeing him on that hall.

In any case, I want to recognize Congressman Stearns, because we will be cochairs of this hearing today. My part will be opening statements and the testimony of the witnesses, an Congressman Stearns will preside during the question-and-answer period after that.

Let me recognize myself for an opening statement.

Today the Subcommittee on Courts, the Internet, and Intellectual Property and the Subcommittee on Commerce, Trade, and Consumer Protection will consider the discussion draft of the "Database and Collections of Information Misappropriation Act.

I am sure many of you are wondering what can be said about database protection that hasn't already been said? However, after 8 years of debate, we are here to review draft legislation that embodies a compromise between the House Committees on the Judiciary and Energy and Commerce.

Electronic compilations and other collections of factual material are absolutely indispensable to the American economy. These information products place a wealth of data at the fingertips of business, professionals, scientists, scholars and consumers. Databases are essential tools for improving productivity, advancing education and training and creating a more informed citizenry.

Developing, compiling, distributing and maintaining databases requires substantial investments of time, personnel and money.

Information companies must dedicate resources to gathering and verifying factual material, presenting it in a user friendly way and keeping it current.

U.S. firms have been the world leaders in this field, but several recent legal and technological developments threaten to erode incentives for investments needed to maintain and expand databases.

While the 1991 Supreme Court decision in Feist Publications reaffirmed that most commercially significant databases satisfy the originality requirement for protection under copyright, the court emphasized that this protection is necessarily thin.

Several subsequent lower court decisions have pointed out that current copyright laws cannot stop a competitor from lifting massive amounts of factual material from a copyrighted publication to use as a basis for its own competing product.

In cyberspace, technological developments represent a threat as well as an opportunity for collections of information. Copying factual material from a third party's collection and rearranging it to form a competing information product is cheaper and easier than ever.

The draft legislation before us today provides protection to databases and gives incentives to their creators to continue producing these invaluable tools. This legislation is a compromise. In fact, a key element is the misappropriation approach that is narrowly tailored to target bad actors while preserving the ability of consumers to access and use information.

Mr. SMITH. Now, that concludes my opening statement, and the gentleman from Florida, Mr. Stearns, is recognized for his.

Mr. STEARNS. Thank you, Chairman Smith, for hosting this unprecedented joint hearing, and on behalf of my fellow Energy and Commerce Subcommittee Members, I want to thank you for the warm hospitality. We are delighted to be here.

The copyright clause of the United States Constitution states that "Congress shall have the power to promote the progress of science and useful arts by securing, for limited times to authors, the exclusive right to their respective writings." .

This power is limited by subject matter, only writings and discoveries of authors may be protected, purposed material may only be protected to the end of promoting science and useful arts; and duration, writings may be protected for a limited time period.

While all three limitations are important, it is the subject matter limitation that is the central consideration underlying copyright protection.

For 7 years, there was a split in the courts about whether copyright protection would be afforded to only creative works or whether noncreative compilations of information could receive protection.

A minority of courts held that noncreative compilations of information could receive copyright protection under the judicial "sweat of the brow" doctrine.

In 1991 the Supreme Court struck down the "sweat of the brow" doctrine. The court wrestled with the ostensible paradox that while facts are not copyrightable, compilation of facts generally are.

The Supreme Court explained that the key to understanding the seeming paradox was in understanding why facts are not copyrightable.

The court held that originality is the sine qua non of copyright law, without regard to the resources spent in collecting and assembling factual compilations.

A compilation is no more worthy of copyright protection than the underlying facts themselves, unless there is a modicum of creativity in the compilation. In other words, creative compilations of information would be protected by copyright. Noncreative compilations of information like the White Pages would not.

And that is why we are here today. Proponents of legislation argue that this decision left a gaping hole in the protection of their products. They believe the distribution capabilities of the Internet have exacerbated the need to fill this gap in protection. Opponents of the legislation see no shortcomings in the current law and believe that proponents of legislation have failed to demonstrate a concrete problem that requires a legislative solution. They believe contract, trespass, misappropriation, unfair competition, and the Computer Fraud and Abuse Act provide sufficient protection for noncreative databases.

It seems that while the opponents of the legislation would support a narrow misappropriation statute, they raise constitutional concerns about broader proprietary interest in factual information, and I share those constitutional concerns.

I believe that Congress should not create property rights in facts. Specifically, I am concerned that the prohibition against making database information available has ambiguous terms that will chill the development of new databases and lead to further litigation. I am concerned that a database that is merely maintained and not necessarily collected would receive protection. How does this standard couple with the time sensitivity standard? Could the maintenance provision cause a court to have a liberal reading of time sensitivity?

I am most concerned about the way this legislation will impact scientific educational and research activities. I worry that the determination of what is customary is so vague, that it will only be resolved through costly litigation. This could put a real chill on important research activity.

As a result, I suspect none of us would like to see this.

And in conclusion, I look forward to a rigorous discussion of the constitutional issues such as the constitutional boundaries of noncreative database misappropriation legislation and the other issues I have raised earlier. I am pleased that we have the opportunity today to have these questions answered. I withhold comment on the draft of the bill until I can be certain that this draft strikes the appropriate balance between access to information, innovation, and protection against misappropriation.

It is plausible that such a balance may be unattainable. I am certain the discussion here today will assist us, and I look forward to hearing from our distinguished panel. And I thank again Chairman Smith for his hosting this joint Committee and his hospitality.

Mr. SMITH. Thank you, Chairman Stearns.

Mr. SMITH. The gentleman from California, Mr. Berman, is recognized for his opening statement.

Mr. BERMAN. Thank you very much, Mr. Chairman, and without, at this point specifically, reacting to some of the comments of my colleague from Florida, I do note the former Chairman of this Subcommittee sitting in the back row there, Howard Coble, and there is something about database protection in Howard Coble that automatically come to mind for anyone who sat through the many hours of hearings and markups of this legislation in earlier Congresses. And we are still with the issue, and it is good to have Howard with us at this time.

I am open-minded on this issue, one of those rare issues that I am open minded on. And I look forward to the witnesses and understanding exactly what the draft does, how it differs from the earlier legislation offered by each Committee and studying the issue further. So I am glad you called this hearing, and I think this is an important issue for us to be dealing with and look forward to the testimony of the witnesses.

Mr. SMITH. Thank you, Mr. Berman.

Mr. SMITH. And I also thank you for pointing out that Mr. Coble is with us, and I would like to ask him, if Mr. Coble is still here, he is welcome to join us up at the table here. And we appreciate all that he has done on this issue to date.

Mr. COBLE. I want to thank the gentleman from California for his kind words. Howard, thank you, but pardon my gravelly voice. I am just getting over a cold. And, Mr. Chairman, thank you for calling—this is—I don't think the gentleman from Florida and I are in synch on this, but this is a very important issue, Mr. Chairman. I thank you for having the hearing.

Mr. SMITH. Thank you, Mr. Coble.

The gentlewoman from Illinois, Ms. Schakowsky is recognized for her opening statement.

Ms. SCHAKOWSKY. Thank you, Chairman Smith, and I want to thank my Chairman, Chairman Stearns, for holding today's hearing on the database and collection of Information Misappropriation Act of 2003.

I really look forward to hearing the expert testimony from today's witnesses, their thoughts on the draft bill and the problems that the bill aims to address.

As we all know, this is a highly controversial issue that has been debated for several years by Members of both Committees. In my view, our copyright laws need to strike a very delicate balance between the interests of proprietors and consumers. Our laws must ensure that proprietors are rewarded for their work, while at the same time protecting the consumer's access to information. It is extremely important that we do not pass laws that prevent people from obtaining factual information. Database owners can already copyright their original selection, coordination and arrangement of facts.

We need to explore the issue further to determine if the current protections are adequate. If they are not, legislative remedies must be narrowly tailored to solve identified problems. Sweeping legislation could hamper research and harm consumers.

I heard just in the last few days from Northwestern University, a University of my district, and they were supporting a letter from the president of the Association of American Universities that was written to Chairman Tauzin and Sensenbrenner that I would—if it hasn't already—like to ask unanimous consent to place the letter in the record.

Mr. SMITH. And without objection, that letter will be made a part of the record.

[The information referred to follows:]

*ACE American Council on Education*
*AAU Association of American Universities*
*NASULGC National Association of State Universities and Land-Grant Colleges*

September 8, 2003

The Honorable F. James Sensenbrenner, Jr.
Chair, House Committee on the Judiciary
2449 Rayburn House Office Building
Washington, DC 20515

The Honorable W.J. " Billy" Tauzin
Chair, House Committee on Energy and Commerce
2183 Rayburn House Office Building
Washington, DC 20515

RECEIVED
SEP 1 0 2003
Committee on the Judiciary

Dear Chairman Sensenbrenner and Chairman Tauzin:

I write on behalf of the Association of American Universities, the American Council on Education, and the National Association of State Universities and Land-Grant Colleges, concerning the database protection legislation discussion draft "Database and Collections of Information Misappropriations Act," which was made available August 29 for consideration by interested parties. Our associations have been involved in the database deliberations from the beginning because of their potentially sweeping implications for the vitality and growth of college and university research and education programs, which invigorate our domestic economy, sharpen our international competitiveness, strengthen our national security, and enrich the health and quality of life of our citizens.

The higher education community has important interests on both sides of the database protection debate. Information, including data of all kinds and databases containing that information, is an essential input into higher education research and education programs. At the same time, databases and organized collections of information constitute one of the most important outputs of those programs. Colleges and universities therefore recognize the need for a balanced approach. We support policies that retain incentives to create databases and that protect the integrity of those databases. But higher education has an overriding interest in preserving unfettered access to data and information, which are, after all, the basic building blocks of the knowledge produced through research and education. Thus, colleges and universities believe that any database protection legislation should provide narrowly focused protection that supports data integrity and responds to demonstrable threats to the incentive to create databases without impeding access to the data and information upon which research and education programs depend.

We appreciate the evident hard work and good-faith effort to achieve an appropriate balance that are reflected in the discussion draft. We understand the difficulty of reconciling the often sharply differing views of the wide range of groups critically affected by database protection legislation.

American Council on Education • One Dupont Circle, NW, Suite 800, Washington, DC 20036 • (202) 939-9300
Association of American Universities • 1200 New York Ave., NW, Suite 550, Washington, DC 20005 • (202) 408-7500
National Association of State Universities and Land-Grant Colleges • 1307 New York Ave., NW, Suite 400, Washington, DC 20005 • (202) 478-6040

From our perspective, we believe that the draft moves in the right direction in a number of important respects when compared to previous legislative proposals. Notably, we recognize and appreciate the elimination of the concept of "qualitative" substantiality, which we believed was inappropriate in a bill intended to protect the substantial effort involved in creating large databases. We also appreciate (i) the clear recognition that the protection of the bill is directed to "large" databases that require substantial effort to develop, (ii) the effort to tie the requisite injury to direct competition in the same market as the existing database, and (iii) the insertion of a knowledge standard as a condition of liability, which should help protect innocent downstream users of data.

However, we are concerned that a number of the problematic provisions and uncertainties of earlier proposals remain in the current draft. For example, the draft does not make clear that making available a "quantitatively substantial part" of a database "generated...through a substantial expenditure..." is not a violation unless the part made available was itself the result of a "substantial expenditure..." by the database claimant. In addition, there appears to be no requirement that the defendant's conduct cause the substantial injury that undermines the incentive to create the database. While one provision includes a related concept, that provision appears to require only proof of the notion that free-riding in the abstract, not necessarily connected to the behavior of the defendant, would diminish incentive, a proposition that, as specified in the bill, would virtually always be true. Thus, the bill would still appear to permit suit for single lost sale or, equally troubling, a single lost source of data. Further, while we understand the threat posed by the unauthorized making all or substantially all of a database available to the relevant public, the concept of "making available to others" appears overly broad, and could pose a threat to traditional, collaborative work within or among universities. We also are puzzled by the apparent lack of any required nexus between the person who contributes the substantial effort or investment that qualifies for protection and the rights and remedies that are granted to any "injured" person.

We are concerned that the bill does not make sufficiently clear that customary academic uses of data and information are protected from liability so that scholars and researchers who engage in those uses will not need lawyers looking over their shoulders lest they find themselves in court facing potential large liability. Protecting academic uses of data has been a fundamental concern of the higher education community since the database discussions began. We believe there should be a strong safety net to prevent the chilling of customary and well-accepted educational and research activities. The lack of such protection could slow the advancement of knowledge and could thus be extremely harmful to the national interest. We have urged, and continue to urge, that the burden of proof of demonstrating that customary nonprofit academic and scientific uses are unreasonable should be a heavy one and should fall on the plaintiff.

In addition, the current draft contains a number of new provisions whose intent and impact are ambiguous and which could have serious unintended consequences for colleges and universities and for research and education. We note in particular (i) the new subpoena provision, which appears to be substantially broader than the already controversial provision in the DMCA, (ii) the service provider immunity provision, which could be read not to provide essential protection that we believe was intended, and (iii) the effect of the inclusion of "maintenance" in the prohibition and how it relates to the duration of protection and the extent of effort required for protection.

Given the importance of data and databases to colleges and universities and the potential long-term economic, educational, and research impacts of the legislation, we urge you to allow sufficient time to address and resolve these concerns before your committees move this bill forward.

We recognize that the database debate has been a long and difficult one, and we understand your desire to advance the process. We are prepared to work with you in any way we can to resolve our concerns and help produce database legislation that provides appropriately focused protection while preserving critical access to data and information.

Cordially,

Nils Hasselmo
President
Association of American Universities

cc: Members of House Judiciary Committee
Members of House Energy and Commerce Committee

---

Ms. SCHAKOWSKY. Let me just quote a couple sentences. Quote, colleges and universities believe that any database protection legislation should provide narrowly focused protection that supports data integrity and response to demonstrable threats to the incentive to create databases without impeding access to the data and information upon which research and education programs depend.

We have also heard from librarians who have been very careful in representing the consumer interests and researchers' interest, and I look forward to learning more about this important topic from today's witnesses. Thank you.

Mr. SMITH. Thank you, Ms. Schakowsky.

Mr. SMITH. Without objection, other Members' opening statements will be made a part of the record, and now I will introduce our witnesses today. Our first witness is David Carson, general counsel of the U.S. Copyright Office. Mr. Carson oversees the Office's regulatory activities, litigation and administration of the copyright law. He also serves as a liaison on legal and policy matters between the Copyright Office and Congress and other Government agencies. He is a graduate of Stanford University where he earned a master's degree in history, and Harvard Law School.

The next witness is Thomas J. Donohue, president and chief executive officer of the United States Chamber of Commerce. Prior to his current post, Mr. Donohue served for 13 years as the president and chief executive officer of the American Trucking Association. Mr. Donohue earned a bachelor's degree from St. John's University and his MBA from Adelphi University.

Our next witness is Keith Kupferschmid, vice president for Intellectual Property Policy and Enforcement for the Software and Information Industry Association. Mr. Kupferschmid is responsible for working directly with SIIA's intellectual property committee. He graduated from the University of Rochester in 1987 with a BS in mechanical engineering and from American University's Law School in 1993.

Our last witness is William Wulf, who was elected president of the National Academy of Engineering in 1997. The NAE and National Academy of Sciences operate under Congressional charter to provide advice to the Government on issues of science and engineering.

Mr. Wulf is the author of over 100 papers and technical reports, has written three books and holds two U.S. patents. And I welcome you all. We have written statements from every one of our witnesses, and without objection, the complete statements will be made a part of the record.

Mr. SMITH. As you know, we expect you to limit your testimony to 5 minutes, and Mr. Carson, we will begin with you.

## STATEMENT OF DAVID CARSON, GENERAL COUNSEL, COPYRIGHT OFFICE OF THE UNITED STATES, LIBRARY OF CONGRESS

Mr. CARSON. Good afternoon, Chairman Smith, Chairman Stearns, Ranking Members Berman and Schakowsky. Members of both Subcommittees. Thank you for giving the Copyright Office the opportunity to testify at this hearing on the discussion draft of the Database and Collections of Information Misappropriation Act.

Our written testimony provides some historical perspective on database protection in the United States, briefly reviews the approach taken in the discussion draft and addresses some of the concerns that critics of database legislation have voiced thus far.

In the few minutes that I have with you today, I would like to focus on the message that the Register of Copyrights has delivered in past testimony on database legislation, a message that is equally relevant today. Since the Supreme Court in the Feist case withdrew much of the protection that copyright law had previously offered to databases, the Copyright Office has perceived a need to

provide adequate incentives for the production and dissemination of databases.

We believe the databases are important to our economy and culture, both as a component in the development of electronic commerce and as a tool for facilitating scientific, educational and technological advancement. We have long recommended an approach to database protection based on a misappropriation or unfair competition model rather than an exclusive property rights model.

At the same time, we have been concerned about the risks of overprotection. The free flow of information is essential to the advancement of knowledge, technology and culture, and we support legislation that, while ensuring adequate incentives for investment, would not inhibit access and use for socially beneficial purposes in appropriate circumstances.

The discussion draft represents a continuing evolution toward such an approach. We commend the leadership of those who have worked so hard to produce a draft that adopts this approach and takes into account the needs of producers of databases as well as users and members of the educational, scientific and research communities.

While we have not had sufficient opportunity to study the discussion draft to permit us to offer any defensive views on this particular draft, we believe in general that it represents a major step in the direction of enactment of the type of balanced legislation that the Office has long recommended.

I should point out that our testimony on this issue in the past, as well as today, draws heavily on our 1997 Report on Legal Protection of Databases which contains a wealth of information on the subject, to which I commend you.

Since the Supreme Court's 1991 decision in *Feist*, only a thin layer of copyright protection remains for qualifying databases. In order to qualify, they must exhibit some modicum of creativity in the selection, arrangement or coordination of the data in them.

The protection is thin in that only the creative elements—the selection, coordination and arrangement of data—are protected by copyright. In no case is the data itself copyrightable.

One of the most significant cases in this area since Feist is the 2nd Circuit's 1997 decision in *NBA* v. *Motorola*. It is our understanding that the approach taken in the discussion draft is designed to codify the standards set forth in the *NBA* case. Our written testimony sets forth our analysis of the extent to which the discussion draft succeeds in this effort, and we believe that, by and large, it does succeed.

We do believe that further clarification may be necessary in some instances. For example, to determine whether the approach taken in the discussion draft is intended to codify nothing more than the hot news misappropriation doctrine discussed in the *NBA* case and the landmark 1918 Supreme Court decision in *International News Services* v. *Associated Press*.

While the discussion draft requires that the unauthorized making available occur in a time-sensitive manner, courts would be instructed to consider the temporal value of the information in the database within the context of the industry sector involved in determining whether the time-sensitive requirement has been met.

This suggests that perhaps something beyond hot news would be protected, but further thought should be given to clarifying the circumstances under which such protection should be given.

We do believe that serious consideration should be given to protecting more than hot news, but on the other hand, to the extent that the legislation would go beyond protection of hot news, we are inclined to favor imposing some time limit on the duration of protection for a database.

As always, the Copyright Office stands ready to assist you in your further consideration of this proposal, and I will be pleased to answer any questions you may have.

Mr. SMITH. Thank you, Mr. Carson.

[The prepared statement of Mr. Carson follows:]

PREPARED STATEMENT OF DAVID O. CARSON

Good afternoon. Chairman Smith, Chairman Stearns, Congressman Berman, Congressman Schakowsky, Members of both Subcommittees, it is a pleasure to appear before you today. Thank you for giving the Copyright Office the opportunity to testify at this hearing on the discussion draft of the Database and Collections of Information Misappropriation Act.

The Copyright Office has testified twice in recent years before the Subcommittee on Courts and Intellectual Property on legislation to protect databases. In the 105th and 106th Congresses, the Register of Copyrights testified in connection with the proposed Collections of Information Antipiracy Act. That bill was passed by the House in the 105th Congress but no action was taken in the Senate. In her testimony on that legislation and on a later version, the Register testified that there was a need to preserve adequate incentives for the production and dissemination of databases, which are increasingly important to the U.S. economy and culture, both as a component in the development of electronic commerce and as a tool for facilitating scientific, educational and technological advancement. She stated that there was a gap in existing legal protection, which could not be satisfactorily filled through the use of technology alone. This legal gap was compounded by the ease and speed with which a database can be copied and disseminated, using today's digital and scanning capabilities. Without legislation to fill the gap, publishers were likely to react to the lack of security by investing less in the production of databases, or disseminating them less broadly. The result would be an overall loss to the public of the benefits of access to the information that would otherwise have been made available.

At the same time, the Register cautioned that the risks of over-protection were equally serious, because (as already noted) the free flow of information is essential to the advancement of knowledge, technology and culture. She testified in support of legislation that would ensure adequate incentives for investment, without inhibiting access for appropriate purposes and in appropriate circumstances.

Accordingly, the Register recommended the restoration of the general level of protection provided in the past under copyright "sweat of the brow" theories, but under a suitable constitutional power, with flexibility built in for uses in the public interest in a manner similar to the function played by fair use in copyright law. Such balanced legislation could optimize the availability of reliable information to the public.

In the intervening years, nothing has occurred to change the views of the Copyright Office. We continue to believe that balanced legislation should be enacted that would provide appropriate levels of protection for producers of databases, without unnecessarily impeding the free flow of knowledge and information.

The discussion draft represents a continuing evolution of the legislation addressing the protection of databases toward a pure misappropriation approach. In our previous testimony we expressed the view that misappropriation is the best approach to this issue and we commend the leadership of all of those who have worked so hard on this issue for their commitment to craft legislation that takes into account the needs of producers of databases as well as users and members of the educational, scientific and research communities. While we have not had sufficient opportunity to study the discussion draft to permit us to offer any definitive views on this particular draft, we believe in general that it represents a major step in the direction of enactment of the type of balanced legislation the Office has long recommended.

Much of what I say today will be based on the research and findings of the Register in her August 1997 Report on Legal Protection for Databases, which was prepared at the request of Senator Hatch, Chairman of the Senate Committee on the Judiciary. We are aware of no major developments since the time of that Report that have significantly altered the landscape with respect to legal protection for databases.

My testimony today will provide a historical perspective concerning the protection of databases in the United States, briefly review the approach taken in the discussion draft and address some of the concerns that critics of database legislation have voiced.

## I. THE HISTORY OF DATABASE PROTECTION IN THE UNITED STATES

In the terminology of the copyright law, a database is a "compilation." The Copyright Act defines a compilation as "a work formed by the collection and assembling of preexisting materials or of data . . ."[1] Compilations were protected as "books" as early as the Copyright Act of 1790.

Over the course of the nineteenth century, two rationales developed for protecting compilations under copyright. The earliest cases identified the compiler's effort—"his own expense, or skill, or labor, or money"[2]—as the critical contribution justifying protection. This type of analysis came to be known as the "sweat of the brow" doctrine. Analyses under sweat of the brow emphasized both the compilers' efforts and the copiers' "unfair use of the copyrighted work, in order to save themselves the time and labor of original investigation."[3]

During the late nineteenth century, courts began to articulate another basis for copyright protection that generally differed from the labor/investment approach taken in cases involving compilations. In a series of decisions from 1879 to 1903, the Supreme Court held that the "writings" that could be protected under the Copyright Clause of the Constitution included "only such as are original,"[4] and indicated that creativity is a component of originality.[5]

The evolving doctrine of originality was applied by some courts in compilation cases, particularly cases involving compilations of textual materials such as law books. These cases identified the author's critical contribution justifying protection as his judgment in selecting and arranging materials.[6]

This approach coexisted with, rather than supplanted, sweat of the brow cases. Sweat of the brow was applied to cases involving purely factual compilations, such as catalogs and directories.

On the question of the scope of protection afforded to compilations, there was somewhat greater uniformity in the case law. In compilation cases, regardless of the theoretical framework adopted to justify copyright protection, once the plaintiff's work was determined to be copyrightable, courts generally held a defendant to have infringed whenever material was copied from the plaintiff's work. Typically, there was no inquiry as to whether the particular material copied was protected by the plaintiff's copyright. To avoid infringement, a second-comer was required to go to the original sources and compile the material independently, without reference to the earlier work.[7] A common thread running through many of these decisions was the court's desire to prevent the copier from competing unfairly with the compiler by appropriating the fruits of the compiler's efforts or creativity. In this sense, courts treated copyright protection for compilations much like a branch of unfair competition law.

In the Copyright Act of 1976, Congress included in the definition of "compilation" the first express statutory link between compilations and original works of authorship ". . .that are selected, coordinated, or arranged in such a way that the result-

---

[1] 17 U.S.C. § 101.

[2] *Emerson v. Davies*, 8 F. Cas. 615, 619 (C.C.D. Mass. 1845).

[3] *West Pub. Co. v. Lawyers' Co-operative Pub. Co.*, 79 F. 756, 772 (2d Cir. 1897).

[4] *In re* The Trademark Cases, 100 U.S. 82, 94 (1879).

[5] *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239 (1903).

[6] *See e.g., Edward Thompson Co. v. American Lawbook Co.*, 122 F. 922, 924 (2d Cir. 1903)(focusing on "skill and taste of the [plaintiff] in selecting or arranging" materials); *Lawrence v. Dana*, 15 F. Cas. 26, 28, 4 Cliff. 1 (C.C.D. Mass. 1869)("copyright may justly be claimed by an author of a book who has taken existing materials from sources common to all writers, and arranged and combined them in a new form, and given them an application unknown before, for the reason that, in so doing, he has exercised skill and discretion in making the selections, arrangement, and combination . . .").

[7] *See, e.g., Williams v. Smythe*, 110 F. 961 (C.C.M.D. Pa. 1901); *List Publishing Co. v. Keller*, 30 F. 772 (C.C.S.D.N.Y. 1887).

ing work as a whole constitutes a work of authorship."[8] Cases under the 1976 Act were divided about the continuing viability of the sweat of the brow doctrine. Some circuits continued to apply it,[9] while other circuits rejected it, requiring a showing of sufficient creativity in order to entitle a compilation to copyright protection.[10] The Supreme Court resolved the split in the circuits in *Feist Publications, Inc. v. Rural Tel. Serv. Co.*[11] In that case, the Supreme Court held that the white pages of a telephone directory (containing an alphabetical listing of all residents with telephone service in a defined geographic area) was insufficiently creative to merit copyright protection. The Court held that the requirement of creativity was not merely statutory, but rooted in the Copyright Clause itself.[12] Thus, the sweat of the brow doctrine was laid to rest.

What remains is a thin layer of copyright protection for qualifying databases. In order to qualify, they must exhibit some modicum of creativity in the selection, arrangement, or coordination of the data. The protection is thin in that only the creative elements (selection, arrangement, or coordination of data) are protected by copyright. Explanatory materials such as introductions or footnotes to databases may also be copyrightable. But in no case is the data itself (as distinguished from its selection, coordination or arrangement) copyrightable. The absence of uniform protection for noncreative databases is what has given rise to the calls for this legislation.

## II. DISCUSSION DRAFT OF THE DATABASE AND COLLECTIONS OF INFORMATION MISAPPROPRIATION ACT

It is our understanding that the scope and applicability of the prohibitions in the discussion draft are designed to codify the standards set forth in the Second Circuit's decision in *National Basketball Ass'n v. Motorola, Inc. ("NBA")*.[13] That case involved a state law misappropriation claim by the NBA against the maker of a hand-held pager which provided subscribers with scores and statistics of professional basketball games in progress.[14] In analyzing the case, the court concluded that a "hot news" misappropriation claim under the theory of *International News Service v. Associated Press*[15] ("INS") would survive preemption by federal copyright law.[16] The court enumerated five elements "central to an INS claim." Those conditions are:

(i) the plaintiff generates or collects information at some cost or expense;

(ii) the value of the information is highly time-sensitive;

(iii) the defendant's use of the information constitutes free-riding on the plaintiff's costly efforts to generate or collect it;

(iv) the defendant's use of the information is in direct competition with a product or service offered by the plaintiff; and

(v) the ability of other parties to free-ride on the efforts of the plaintiff would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened.[17]

### *A. The plaintiff generates or collects information at some cost or expense.*

The first condition is codified in subsection 3(a)(1) of the discussion draft, which applies the prohibition against misappropriation only to databases that were "generated, gathered, or maintained through a substantial expenditure of financial resources or time." The term "maintained" does not appear in the court's articulation

---

[8] 17 U.S.C. § 101.

[9] *See, e.g., Illinois Bell Tel. Co. v. Haines & Co.*, 683 F. Supp. 1204 (N.D. Ill. 1988), *aff'd*, 905 F.2d 1081 (7th Cir. 1990), *vacated and remanded*, 499 U.S. 944 (1991); *Rural Tel. Serv. Co. v. Feist Publications, Inc.*, 916 F.2d 718 (10th Cir. 1990), *reversed*, 499 U.S. 340 (1991).

[10] *See, e.g., Financial Info., Inc. v. Moody's Investors Serv., Inc.*, 808 F.2d 204 (2d Cir. 1986), *cert denied*, 484 U.S. 820 (1987); *Eckes v. Card Prices Update*, 736 F.2d 859 (2d Cir. 1984); *Worth v. Selchow & Richter Co.*, 827 F.2d 569, 572–73 (9th Cir. 1987).

[11] 499 U.S. 340 (1991).

[12] *Id.* at 346.

[13] 105 F.3d 841 (2d Cir. 1997).

[14] The case also involved a claim of infringement of the copyrights in the broadcasts of the games. That claim was rejected by the court because alleged infringement involved reproduction only of the uncopyrightable facts from the broadcasts, and not of the expression or descriptions of the games that constituted the broadcasts. 105 F.3d at 847.

[15] 248 U.S. 215 (1918).

[16] *See* 17 U.S.C. § 301.

[17] 105 F.3d at 852.

of the first condition. However, the reference to "quality" in the fifth factor could suggest a recognition that misappropriation applies not only to the initial creation but to the periodic update and verification of the product or service. One other variation from the exact language of the court is the requirement of a "substantial" expense. The court used the arguably less demanding term, "some." The discussion draft also equates "time" with "cost or expense," which we believe is probably a defensible interpretation of the elements set forth in NBA.

*B. The value of the information is highly time-sensitive.*

The second condition is codified in subsection 3(a)(1) of the discussion draft, which requires that the making available occur "in a time sensitive manner." Section 3(c) of the discussion draft states that courts shall consider "the temporal value of the information in the database, within the context of the industry sector involved" in determining whether this condition is met. The discussion draft omits the term "highly," although it is not clear how much difference that makes. The discussion draft appears to take a flexible approach to this condition, requiring consideration of the business context, but also allowing a court to consider whatever other factors it might deem relevant. This approach may well be the subject of initial uncertainty, until courts have provided guidance in applying the standard. In this respect, the discussion draft may go beyond the "hot news" doctrine addressed in *NBA* and *INS*.

In its previous testimony, the Copyright Office noted with approval the application of a definite term of protection, beginning at the time the relevant portion of the collection is first used in commerce. The Office continues to have concerns about protection without a clear end point. However, the time sensitivity provisions of the discussion draft may address that concern, depending upon how they are interpreted. It may be that consideration should be given to clarifying the scope and application of the "time sensitive" component of this discussion draft. To the extent that it goes beyond "hot news"—and in the past the Office has supported protecting more than "hot news"—there may still be reason to consider some specific limitation on the duration of protection.

*C. The defendant's use of the information constitutes free-riding on the plaintiff's costly efforts to generate or collect it.*

The third condition is codified in subsection 3(a) of the discussion draft, which prohibits the "mak[ing] available in commerce to others a quantitatively substantial part of the information in a database generated, gathered, or maintained by another person . . ." While the term "free- riding" does not appear in the relevant portion of the text, the conditions described appear to be the practical equivalent. Moreover, the "free-riding" problem is addressed in subsection 3(a)(3).

*D. The defendant's use of the information is in direct competition with a product or service offered by the plaintiff.*

The fourth condition is codified in subsection 3(a)(2) of the discussion draft, which requires that the making available "inflict[] an injury." That term is defined in subsection 3(b) as "serving as a functional equivalent in the same market as the database in a manner that causes the displacement, or the disruption of the sources, of sales, licenses, advertising, or other revenue." Here the discussion draft expressly provides for direct competition and also requires the showing of at least some disruption in revenue to the compiler.

*E. The ability of other parties to free-ride on the efforts of the plaintiff would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened.*

The fifth condition is codified verbatim in subsection 3(a)(3) of the discussion draft. Thus, this legislation appears to codify the standards set forth by the Second Circuit.

## III. CRITICISMS OF THE DISCUSSION DRAFT

I understand that the discussion draft has been the subject of criticism. I would like to take this opportunity to address some of those arguments.

*A. Constitutionality*

It has been suggested that this legislation exceeds Congress' authority under Article I, section 8 of the Constitution. As you know, the Constitution provides explicit

authority for the protection of copyright.[18] As discussed earlier, the Supreme Court held in *Feist* that the Copyright Clause cannot serve as a basis of authority for the protection of noncreative databases. But *Feist* does not address whether some other basis for protection of such materials may exist. The most likely other basis is the Commerce Clause.[19] At least one critic suggests that the Commerce Clause cannot serve this function. The Copyright Office disagrees.

It has long been accepted that Congress has the power to enact trademark legislation under the Commerce Clause, despite the fact that trademarks may be seen as a form of intellectual property; that trademark law protects material that does not meet standards for copyright and patent protection; and that the protection may last indefinitely. The Supreme Court's opinion in *The Trademark Cases*[20] held unconstitutional an early attempt by Congress to enact a trademark law, based on a lack of Congressional power under either the Copyright Clause or the Commerce Clause. According to the Court, the Copyright Clause did not provide authority for the legislation because trademarks have different "essential characteristics" from inventions or writings, since they are the result of use (often of already-existing material) rather than invention or creation, and do not depend on novelty or originality.[21] The Commerce Clause did not provide authority because the particular trademark law in question governed all commerce and was not limited to interstate or foreign commerce.[22] The opinion suggested that similar legislation limited as to the type of commerce involved would pass constitutional muster under the Commerce Clause. Indeed, legislation consistent with the Court's interpretation of the Commerce Clause was subsequently enacted and has gone unchallenged since 1905.

The Register's 1997 Report on the Legal Protection of Databases stated that "To the extent that database protection promotes different policies from copyright protection, and does so in a different manner, it is similar to trademark law, and therefore seems likely to survive a constitutional challenge." The prohibition set forth in this discussion draft appears to meet that prescription. It is crafted to protect that which the *NBA* case held to be outside of copyright. Its focus is on unfair competition through the misappropriation of a commercial product that is the result of substantial expenditure of another's financial resources or time, in a way that inflicts commercial injury on that person, elements that are far removed from the core of copyright.

*B. Subpoena to Identify Violators*

The discussion draft includes a procedure similar to that in 17 U.S.C. § 512(h) to allow potential plaintiffs to learn the identity of those they believe have violated the provisions in this discussion draft. The Copyright Office believes that the section 512(h) subpoenas are a necessary and appropriate tool in copyright owners' struggle against infringement, particularly in the digital and online environments. However the discussion draft does differ in one significant respect: Section 512(h) requires the person seeking a subpoena to file with the clerk of the court a certain information about the claim of infringement that has given rise to the controversy that requires identification of the alleged infringer. This provision provides assurances that the subpoena is sought in good faith and that there is an objective basis for seeking it. The current discussion draft does not have any analogous safeguards. The Copyright Office recommends the inclusion of such a provision in this discussion draft.

*C. Fair Use Exception*

We understand that some have suggested that this discussion draft is somehow flawed without the inclusion of a "fair use" exception, similar to the one that appears in the Copyright Act.[23] In the past, the Copyright Office has supported inclusion of provisions similar to fair use in database protection legislation. However, the past legislative proposals provided for broader protection than is provided in this discussion draft. In providing for a narrower prohibition, the discussion draft may well obviate the need for a fair use-type of provision. It may well be that this discussion draft already incorporates most of the principles embodied in copyright fair use.

---

[18] "To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." U.S. Const., Art. I, sec. 8, cl. 8.

[19] "To regulate Commerce with foreign Nations, and among the several States . . ." U.S. Const., Art. I, sec 8, cl. 3.

[20] 100 U.S. 82 (1879).

[21] *Id.* at 93–94.

[22] *Id.* at 97.

[23] See 17 U.S.C. § 107.

The "purpose and character of the use"[24] is addressed by subsection 3(a) of the discussion draft, which prohibits the "making available in commerce," and in subsection 3(b), which makes clear that the prohibition extends only to inflictions of injury that serve as a functional equivalent in the same market as the database. The "amount and substantiality of the portion used"[25] is also addressed in subsection 3(a), which requires "a quantitatively substantial part of the information." Indeed, this provision is more permissive than fair use, which may not excuse the use of a quantitatively insubstantial portion that is qualitatively vital to the work. The "effect of the use upon the potential market for or value of the copyrighted work"[26] is addressed by subsection 3(a)(3), requiring that the ability of others to free-ride threaten the "existence or quality" of the database, as well as subsection 3(b), with its strong requirement of market harm. Of course, the second fair use factor, "the nature of the copyrighted work,"[27] is inapplicable to a legal regime specifically designed to protect that which is denied copyright protection for lack of creativity. While we are strong proponents of fair use and understand the desire for such a provision in database protection legislation, we are not persuaded that such a provision is necessarily required when the prohibition itself serves the policies underlying fair use.

*D. Internet Service Provider Liability*

There has been complaint that the discussion draft would subject internet service providers ("ISPs") to liability unfairly. However, subsection 7(i) of the discussion draft explicitly insulates ISPs from liability unless their employees violate the prohibition while acting within the scope of their duties, actively direct or induce a violation of the prohibition, or receive a financial gain directly attributable to the violative conduct. It is not readily obvious to the Copyright Office how the ordinary use of ISPs' systems by their users could be within the scope of these few exceptions to the general rule that ISPs do not bear liability under this discussion draft. Moreover, it is notable that the discussion draft provides this benefit to ISPs without requiring them to abide by many of the conditions that appear in section 512 of the Copyright Act,[28] such as taking down violative material in response to a notice or terminating the account of a repeat offender. Compared to section 512, this provision appears to be generous.

*E. Alleged Expansion of Intellectual Property Protection*

There is also apparently a somewhat amorphous criticism that this discussion draft would serve in furtherance of an alleged trend of expanding intellectual property protection without counterbalancing other interests. The Copyright Office sees no such trend. Indeed, the last few years have seen expansions of exceptions and limitations. For example, legislation has provided exceptions and limitations for digital distance education,[29] use of works by the blind,[30] and the aforementioned provisions for ISPs.[31]

A complete analysis of intellectual property protection includes a consideration not only of the provisions of the law, but also of the other factors which affect the incentive to create and the availability for use of protected materials. Most significantly, the dramatic growth of the use of digital technology and the Internet have made more materials available to more people than ever before. However, this technology has also created an avenue for the improper use of materials on a previously unimagined scale. Changes in the law to try to prevent or remedy these improper uses do not necessarily reflect a change in philosophy about the appropriate scope of protection and have not altered the fact that both authorized and unauthorized users of protected materials generally have greater opportunities to use the material of others than they did before these technological developments.

## IV. CONCLUSION

The discussion draft represents the latest in a series of legislative attempts to provide consistent, federal standards of protection for databases. As I noted at the out-

[24] § 107(1).
[25] § 107(3).
[26] § 107(4).
[27] § 107(2).
[28] *See* 17 U.S.C. § 512 (limiting the liability of qualifying ISPs for copyright infringement).
[29] *See* 17 U.S.C. § 110(2).
[30] *See* 17 U.S.C. § 121.
[31] *See* 17 U.S.C. § 512.

set, the Copyright Office is sympathetic to these efforts but does not, at this time, take a position on this legislation. As always, the Copyright Office stands ready to assist both Subcommittees and I will be pleased to answer any questions you may have.

Mr. SMITH. Mr. Donohue.

**STATEMENT OF THOMAS J. DONOHUE, PRESIDENT AND CHIEF EXECUTIVE OFFICER, CHAMBER OF COMMERCE**

Mr. DONOHUE. Thank you, Mr. Chairman. And Mr. Chairmans, all and Members, it is fun to appear before a joint Committee. You are never quite sure where the balance is. So we will keep an eye from this end as well.

On behalf of the Chamber's 3 million business members of every size and every sector, I am pleased to come here today and to discuss this draft legislation. Let me briefly explain the two very important reasons why the Chamber is clearly opposed to this piece of legislation, and then we will gladly answer your questions.

First, the Nation already has on the books, as many have already mentioned, the vast web of laws that protect database information, laws that even database producers themselves appear to be completely satisfied with. We have the benefit of contract intellectual property, copyrights, state of law misappropriation, trespass and Federal computer antihijacking statutes and numerous other protections that are on the books and in place.

The Chamber's members, you will not be surprised, include many of the country's biggest producers and users of databases, and I would clearly tell you if they didn't think the law was protecting them because they would have clearly told me; but in fact, our Members have told us that this legislation is misdirected, in many ways, harmful and unnecessary.

Proponents of the database legislation have yet to provide a real-world example of a database that isn't protected under current law. I agree under the discussion that the law has been strong and weak in various ends of the bookends, but clearly there is protection; and if we were to pass this draft legislation, there would be many new problems for all of us to face.

It is remarkable that when you think about the enormous number of databases that our children use and we use and that all your staffs use every day, and the equally enormous number of opportunities for some kind of serious infringement is that no one here is telling you about one that has happened. Six years ago the Copyright Office told the Congress that they could pass legislation to add additional protection. The Congress has not, during that period of time, with the introduction of new technology, with people carrying Blackberries around on their belt so they can access all kinds of information, we haven't had these problems.

This is a solution in search of a problem, and we ought to be very careful about that.

Now, your mission in the Congress—the Congress has many missions, but in this instance it is to specifically identify and define the problem and then craft some legislation to fix it. I think you will have to spend the majority of your time looking for the problem.

Let me go to the second concern I have, which is even more critical, and that is this legislation with its vague terms and allowance for excessive penalties—and I will say something about that at the end—would significantly add to the country's frivolous litigation nightmare.

For example, if this legislation were enacted, an individual or an organization such as the Chamber could be sued for taking the text of existing laws and reformatting them to make them easier for people to understand. It could also stop a replacements part manufacturer from being able to compare and hence advertise their products as alternatives to other more expensive originals; and with the legislation's vague and expansive definition of the term time-sensitive, information—time-sensitive information, an individual or an organization could be the target of a lawsuit for using information in a database that is decades old.

Ladies and gentlemen, I think for a minute of where our society would be without sharing database information. We owe an incredible advances in medicine, science, technology and the arts to the availability of that information to researchers all over the place.

That is not to say, however, that access to database information should necessarily be available free. When people reformat and lay it out in a usable and a helpful way, they can and should charge for it, and the user should pay for it.

Passing this legislation—let me conclude—would put a chill on business investment, deprive consumers of new information products and threaten a litigation bonanza that we can't afford, and it would then put penalties that will quadruple the fines when the things we do in RICO statutes only triple them. This would be a very unfortunate piece of legislation.

Thank you very much, Mr. Chairman.

Mr. SMITH. Thank you, Mr. Donohue.

[The prepared statement of Mr. Donohue follows:]

PREPARED STATEMENT OF MR. THOMAS J. DONOHUE

INTRODUCTION

Chairman Stearns, Chairman Smith, Ranking Member Schakowsky, Ranking Member Berman and Members of the Committees. Thank you for the opportunity to be here today to testify on "the Database and Collections of Information Misappropriation Act".

I'm Tom Donohue, President and Chief Executive Officer of the U.S. Chamber of Commerce, the world's largest business federation, representing more than 3 million businesses of every size, sector and region of the country.

I have previously written to both Committees to express my serious concerns about the draft legislation. I ask that copies of my correspondence be made a part of the record.

I recognize that the staffs of your two Committees have worked hard to address the concerns that the Chamber and others have expressed regarding this issue. Unfortunately, I believe that the draft Database and Collections of Information Misappropriation Act does not adequately address those concerns. Instead, this legislation continues to pose a serious threat to the business community, as well as to the academic and science community. I strongly urge you not to move forward with this legislation.

Although the discussion draft has been shared with the public for barely three weeks, it has already attracted a firestorm of thoughtful criticism. Along with the Chamber's objections, additional detailed and persuasive criticisms have been lodged by an extraordinarily diverse array of public and private entities. When the U.S. Chamber of Commerce, Consumers Union, the American Conservative Union, Association of Research Libraries NetCoalition and the Eagle Forum, all join hands in

opposition to a legislative initiative, it's fair to ask what could unite groups which so often see the world in very different ways.

The reason why so many organizations are so concerned about the discussion draft is simple. There has been no threshold showing that there is a problem that needs to be addressed by legislation. However, there is enormous concern that were this draft enacted, it may well create enormous problems for information users and producers, stifling innovation and adding to the excessive litigation burdens already facing American businesses.

We live in the "Information Age" - an age in which advances in information technology have helped fuel economic growth and enhanced productivity. Fundamental changes in basic information policy will affect virtually every American, as well as virtually every business, not just those commonly thought of as information companies.

Our country's basic information policy provides that facts - the building blocks of information - cannot be owned. That historic policy was underscored in a unanimous 1991 Supreme Court decision *Feist* v. *Rural Telephone*. In that landmark case, the Supreme Court reaffirmed that the U.S. Constitution prohibits copyright protections for facts contained in a database. The Court concluded that the Constitution's objective of promoting "the Progress of Science and useful arts" is accomplished by "encouraging others to build freely upon the ideas and information conveyed by a work."

The basic goal of copyright, indeed of all intellectual property law, is to encourage creative activity. That is why the *Feist* court underscored that intellectual property protection can only be provided to those portions of a database that reflect a minimal level of creativity. Notwithstanding this unanimous Supreme Court decision, however, proponents of broad database legislation continue to seek protection for information that they haven't created.

That is not to say that access to these databases should necessarily be available for free. Indeed, the Chamber strongly believes that current protections, such as appropriate intellectual property protections, along with contract and licensing agreements and state trespass and misappropriation protections, as well as other protections, should be utilized and enforced. Companies need and deserve protection for the time, effort and expense that they undertake to create databases, but new intellectual property protections like those envisioned by the discussion draft are too broad and unnecessary.

WHAT'S THE PROBLEM? WHERE IS THE "GAP" IN CURRENT LAW?

The Chamber has always believed that the best way to legislate is to specifically identify and define problems, and then carefully craft legislation to deal with those particular real-world harms. Whenever Congress legislates - especially in an area with the broad ramifications such as those inherent in changing basic and longstanding copyright and information policy - that legislation should be narrowly targeted to resolving a demonstrated real-world harm, with as little collateral damage as possible.

In this arena, throughout the seven-year consideration of this issue, proponents of changing how our nation regulates information have yet to provide a real-world example of a database that can't be protected under current law. There are an astronomical number of opportunities daily for some kind of infringement. Yet the inability to cite gaps in the law is profoundly telling. Indeed, this inability to cite real gaps in existing laws underscores our concern that some proponents of broad database legislation seek to leverage dominance in existing markets into dominance in other markets - without having to gain these advantages via competition in the marketplace.

Most persuasive to me is the reaction of the Chamber's members to the discussion draft. Our broad membership includes many of America's most significant database producers. These companies invest enormous sums of money producing creative, new information products. These companies currently enjoy myriad legal protections for their databases, including contract, copyright, state- law misappropriation, trespass, federal computer anti-hacking statutes and numerous other protections.

If our Chamber members believed for a second that they couldn't protect their substantial investments in database production, they would be urging me to affirmatively fight for new law. Instead, I'm hearing that there is little or no upside for the business community in database legislation, and potentially a significant, anticompetitive downside.

THREAT OF EXCESSIVE LITIGATION

As you know, the Chamber has long been concerned about the threat excessive litigation poses to the economy and American business. This legislation, if enacted, would combine vague terms and excessive penalties to create a frivolous litigation nightmare for businesses of all industries.

The reason for this begins with the core prohibition of the draft bill. Since the Supreme Court's 1918 decision in *International News* v. *Associated Press*, courts have awarded relief in what became known as "hot news" misappropriation cases. That line of cases established the tort of misappropriation, and found that even factual data could be protected if the data met a series of tests, including that the data is "highly time sensitive". The Court found in that case that wire stories were "hot" and protected for a few hours. Subsequent cases have found, for example, that sports scores are "hot" and potentially protected for a matter of minutes.

The discussion draft, however, creates a new definition of "time sensitivity" in the context of this bill, significantly different than the "time sensitivity" that courts have been familiar with for more than eighty years under the *International News* line of cases. Specifically, this draft would potentially require courts to add the concept of "value" to the determination of time sensitivity.

For example, this draft legislation works retroactively, ensnaring facts in databases that are conceivably decades old. The draft protects facts in encyclopedias, even though the lead-time in publishing means that data is generally months old before it reaches the bookstores. In short, it is impossible to state definitively what this core prohibition means - though it can be definitely stated that this prohibition bears only a superficial resemblance to the time-sensitivity standard created by the Supreme Court in the *International News* case and expressly preserved in *Feist*.

The courts would be forced to determine whether the proposed prohibition can be tightened to look like constitutionally sanctioned "hot news" misappropriation and not like the copyright of facts forbidden by *Feist*. While the courts sort this out, the combination of vague terms, a private right of action, quadruple damages and incredibly expansive subpoena power would create a litigation bonanza that will chill investment and threaten business, depriving consumers of new information products.

CONCLUSION

On behalf of the Chamber, I want to thank you for the opportunity to share some of our more serious concerns regarding the discussion draft. The Chamber has always believed that the best way to legislate is to identify and define specific problems, and then carefully craft legislation to deal with them. While some urge "moving beyond" discussing the problem in order to legislate, we are convinced that, if there is to be legislation it should be narrowly targeted to resolving a demonstrated real-world harm, with as little collateral damage as possible.

Appropriate information policy is critical to American business. While we may be willing to support compromise legislation carefully targeted to deal with specific, demonstrated "gaps" in existing law, there has been no demonstrated need for such legislation at this time.

On behalf of American businesses and our three million members, I want to thank you again for inviting me to testify and share our concerns.

Mr. SMITH. Mr. Kupferschmid.

## STATEMENT OF KEITH KUPFERSCHMID, VICE PRESIDENT, INTELLECTUAL PROPERTY POLICY AND ENFORCEMENT SOFTWARE AND INFORMATION INDUSTRY ASSOCIATION, ON BEHALF OF THE COALITION AGAINST DATABASE PIRACY

Mr. KUPFERSCHMID. Chairman Smith and Stearns and Members of both Subcommittees, I appreciate the opportunity to testify before you today to discuss the need for legislation that protects America's databases from piracy. I also would like to especially thank Chairman Sensenbrenner and Chairman Tauzin for their strong leadership on this important issue and appreciate the commitment of the two Committees to work together to produce and enact meaningful database legislation.

I am Keith Kupferschmid, vice president of intellectual property for the Software and Information Industry Association, and I am here today on behalf of the coalition against database piracy. CADP is a broad-based coalition that was formed for the sole purpose of pursuing enactment of a Federal law to prevent misappropriation of databases. Its members include large and small database producers who devote substantial resources to creating and distributing database products and services.

The value of reliable and comprehensive databases that these companies make available to researchers, to businesses, to Government officials, to citizens is immeasurable. Farmers use databases to get weather and soil information; lawyers to rely on legal precedent; doctors to determine safe and effective medical procedures; workers to search for new jobs; pharmacists to understand drug interactions; home buyers to find the right house, and the list goes on.

Database piracy is a major concern to America's database publishers. With the Internet and advances in technologies, databases can be easily stolen and made available to others in ways that cause great harm to the original database producer.

Unfortunately, U.S. copyright law and other existing laws do not adequately protect against such piracy. A recent case, Schoolhouse versus Anderson, decided in 2002, demonstrates the glaring inadequacies of current law. In that case the defendant copied and posted on the Internet a minimum of 74 percent of a small magazine publisher's database of school information. Although the defendant admitted to copying the database, the court held that the defendant was not liable for copyright infringement.

Shortly after that case, the plaintiffs in the case got out of the database business. This is just one example. There are many other cases, including *Skinder-Strauss* v. *MCLE*, *EPM Communications* v. *Notara Warren Publishing* v. *Microdos*, *Ticketmaster* v. *Tickets.com* and many others.

In addition to the database piracy cases that have resulted in litigation, there are numerous other instances of piracy that never make it to the courtroom. Many database producers are simply unwilling to spend the significant amounts of money litigating questionable causes of action and in the process draw attention to the vulnerabilities of their company's databases.

Clearly there is a definite and significant need for database protection legislation. In addition, the risk of potential future instances of database piracy and the adverse effects that piracy would have on investment in databases and consumer protection is certainly sufficient justification for Congress to enact database protection legislation.

I would like to focus the remainder of my remarks on the draft legislation. The discussion draft reflects years of discussions and negotiations between the two Committee staffs and stakeholders. The draft legislation they have developed takes a very targeted and very narrow approach to addressing the problem of database piracy.

It is based on a misappropriation approach that only covers the act of making a database available that causes significant commer-

cial harm to the database producer. It protects the database itself, not the information or the facts in the database.

The draft legislation creates a narrowly focused prohibition that applies only if ten criteria are met. These ten criteria, all of which must be satisfied, set a very high standard for qualifying for protection under this draft bill. This standard is even higher when one also considers the exceptions to liability that are contained in the draft.

We believe that some of the substantive provisions of the draft will provide protection against database piracy while also accounting for the legitimate concerns of database users.

However, we believe that some of the language contained in the draft requires clarification, notably the preemption and time sensitivity provisions, among others. We are also concerned that the discussion draft does not recognize database thefts that cause noncompetitive harms. We look forward to working with the two Committees to ensure that these concerns are addressed.

We note that a few groups, many of whom were part of the process initiated by the two Committees to come up with a compromised text, have voiced their opposition to the discussion draft. The approach of the discussion draft, relying on a standard of misappropriation, is precisely the standard that was recommended by many of those who are now writing to express their concern. Their continued opposition amply demonstrates that they simply do not accept the conclusions that the chairmen have both reached: that Congress should legislate to improve legal protection available for databases.

Our goal throughout this whole process has been to get narrowly targeted legislation that will address the problem of database piracy while also addressing the legitimate concerns of the database user community. To the extent that the opponents believe that the draft falls short of this goal, we continue to stand ready to address those concerns in exchange for their support for this important piece of legislation and their recognition that the bill must address the needs of the database publishing community.

We look forward to working with the Congress and the other stakeholders to achieve a legislative solution that eliminates the unfairness we discuss today. Thank you again for all your work on this important legislation. I will be happy to answer any questions.

Mr. SMITH. Thank you, Mr. Kupferschmid.

[The prepared statement of Mr. Kupferschmid follows:]

Statement of

Keith Kupferschmid
Vice President,
Intellectual Property Policy and Enforcement
Software & Information Industry Association (SIIA)

On Behalf of the
Coalition Against Database Piracy

Before the

Subcommittee on Courts, the Internet and Intellectual Property
of the Committee on Judiciary

and

Subcommittee on Commerce, Trade and Consumer Protection
of the Committee on Energy and Commerce

U.S. House of Representatives

**H.R. ____, THE "DATABASE AND COLLECTIONS OF INFORMATION MISAPPROPRIATION ACT"**

September 23, 2003
2141 Rayburn House Office Building

**Summary**

America's database producers have invested hundreds of millions of dollars to collect, organize and maintain information contained in thousands of databases. These large investments have been made to provide easy-to-use valuable information to users in settings ranging from the general public to business and specialized user communities—information that is important in their work and everyday lives. It is critical that database producers be able to protect their investments and encourage new investments in these important information resources.

Misappropriation of databases threatens the availability of organized, timely, comprehensive information. If investments in databases continue to be destroyed, there will be fewer and fewer people willing to make the investment necessary to create and disseminate these valuable database. Moreover, those database compilers who decide to stay in the database business will be forced to keep locked up the information in their databases to avoid destruction of their investment. Inevitably, this will result in fewer and less reliable databases and thus, less information to fuel the information age.

Despite the acknowledged value provided by America's databases, there is presently a lack of meaningful national legal protection for these databases. Certainly, there are some laws available that provide some small amount of protection to database providers, but more often than not these laws fail to adequately deter or prevent databases from piracy. There are a long list of cases that establish this point.

To be clear, we are seeking a very narrowly targeted approach to this problem. We are not seeking "copyright plus," to expand copyright law, to acquire exclusive rights in the database or to lock up information. We are merely trying to protect from free-riders taking our databases and making them available in a way that hurts our businesses. We think this is a reasonable request.

We believe that some of the substantive provisions of the discussion draft, the "Database and Collections of Information Misappropriation Act," will provide protection against database piracy while also accounting for the legitimate concerns of database users through narrowly-crafted exceptions and limitations on liability. Like many other stakeholders, we have concerns with the language used in some of the provisions of the draft. Most significantly, we believe the language in some of the provisions – notably the preemption provisions and time sensitivity provisions, among others -- is somewhat ambiguous and could cause inadvertent consequences. Equally as disconcerting, is that the discussion draft does not recognize database thefts that cause noncompetitive economic harms that adversely affect ISPs and others that have commercial databases. We look forward to working with the Committees to ensure that any preemption of state law is narrowly tailored and does not impede effective licensing of databases or other measures that might otherwise be available, and seeking some appropriate clarifications of the text, including protection against database theft when carried out by or on behalf of parties other than direct competitors.

**Statement Text**

Chairmen Smith and Stearns, Ranking Members Berman and Schakowsky and members of both Subcommittees, I appreciate the opportunity to testify before you today on the need for legislation that adequately and effectively protects America's databases from piracy. I also welcome the opportunity to comment on your discussion draft, titled the "Database and Collections of Information Misappropriation Act," and would like to especially thank Chairman Sensenbrenner and Chairman Tauzin for their strong leadership on this important issue. We appreciate the commitment of the two committees to work together to produce and enact meaningful database protection legislation.

I am Keith Kupferschmid, Vice President of Intellectual Property Policy and Enforcement for the Software & Information Industry Association. I am here today to testify on behalf of the Coalition Against Database Piracy (CADP). CADP is an ad hoc advocacy group that was formed for the sole purpose of pursuing enactment of a federal law to prevent misappropriation of databases. Its members include large and small database producers who devote substantial resources in compiling, organizing, and distributing database products and providing services that rely on databases. A listing of those companies and organizations that support meaningful database protection is attached.

Database protection is a critical issue to America's database publishers. These companies and organizations have invested hundreds of millions of dollars to collect and organize information contained in thousands of databases. Database publishers not only collect,

compile, and organize the information, they also keep it updated and reliable. These investments are worthy and deserving of protection. Investments in these databases have been made to provide easy-to-use valuable information to users in settings ranging from the general public to business and specialized user communities—information that is important in their work and everyday lives. The value of reliable and comprehensive databases that these companies make available to researchers, businesses, government officials, and citizens is immeasurable. Farmers use databases to get weather and soil information; lawyers, to rely on legal precedent; doctors, to determine safe and effective medical procedures; workers, to search for new jobs; pharmacists, to understand drug interactions; home buyers, to find the right house, and the list goes on.

Given the important role that databases play in our capital markets, law enforcement, and science and research, it is critical that database producers be able to protect their investments from free-riders and pernicious commercial exploitation and that new investments in these important information resources be encouraged. Protection of database investments will stimulate the economy by creating incentives for investments in new databases and accelerating job growth in large and small businesses in our nation's vital U.S. information industry. U.S. database and directory publishers were estimated to generate $15.4 billion in annual revenue in 1999.[1] Without effective statutory protection, private firms are deterred from investing in

[1] A few examples of the value of today's databases:

"PoisIndex is an index of approximately one million entries on a wide variety of poisonous substances, including drugs, chemicals, commercial and household products, and biologic substances. Substances are reviewed for entry into the database by a group of skilled medical professionals, who also scan the world's medical literature for pertinent data on toxic exposure and management. Approximately 200 actively practicing clinicians from over 20 countries participate in the editorial and selection process. Each substance entry in the database is linked with up to four full-text documents outlining clinical effects, range of toxicity, treatment measures, and other toxicological information. Software engineers are employed to maintain, test, produce and support the database and the software required to store, edit, sort and retrieve the data. The typical PoisIndex user is a medical professional, usually an

database production, resulting in fewer jobs and a shortfall of reliable, accurate and up-to-date information. Protection will promote investments in the creation of new jobs and information services, increase the pace of technological progress, and fulfill the economy's growth potential.

While databases play a key role in supporting America's information based economy, they are also important in protecting our health and well-being. Databases ensure that our prescription medications are safe and will not adversely interact, provide healthcare professionals with vital information on countless topics such as proper antidotes for poisons, and serve as an important resource for mapping out cures for deadly diseases.

Meaningful database protection legislation will also ensure that consumers and businesses have access to the most accurate and reliable information. While database producers are constantly updating their information, those who pirate databases cannot be relied on to do

---

emergency physician or poison center specialist, who needs instant access to such information in life-threatening circumstances."

"The MDL Drug Data Report (MDDR), produced by MDL Information Systems … is a database of approximately 85,000 chemical compounds with potential drug applications. It is updated on a monthly basis from a specialized search of published reports, patent applications and scientific papers so as to make data available on new biologically active compounds as soon as they are disclosed. MDDR tracks these compounds through stages of development and into clinical trials. Accompanying software permits researchers to analyze the effects of modifications of a drug compound's structure on its properties. Researchers can also combine the results of their own internal and external results with the database supplied by MDL to develop their own specialized research tool."

"Derwent World Patents Index is a comprehensive database of more than seven million separate inventions culled from more than 13 million patent documents worldwide. Coverage includes patents of products from the pharmaceutical industry, agricultural and veterinary medicine, polymers and plastics, chemistry, electronics, electrical and mechanical engineering. All patent information is presented in a uniform, user-friendly format consisting of a simplified English-language abstract explaining key technical details and highlighting applications. In addition to bibliographic information, technical drawings or diagrams are included as available. The Index is updated weekly with information from 40 patent-issuing authorities around the world, 1200 scientific journals, and papers presented at international conferences. Users of the Derwent Index include patent and information professionals, research scientists, engineers, universities, research institutes, libraries, and individual inventors and entrepreneurs." Examples quoted from a report by Laura D'Andrea Tyson and Edward F. Sherry entitled "Statutory Protection for Databases: Economic and Public Policy Issues."

the same. As a result, consumers and businesses may end up relying on outdated information, in some cases to the detriment of their health. For example, a mother trying to find the contraindications for a particular medicine could get the incorrect or out-of-date information by unwittingly accessing a stolen outdated database rather than the original compiler's current database.

Consumers need high quality databases which contribute to their ease, productivity, and innovation. Databases save consumers time and effort. Finding that needle in a haystack from diverse and unorganized sources could take an individual weeks or months. Fortunately, database publishers provide quick and easy answers in their databases.

Misappropriation of databases threatens the availability of organized, timely, comprehensive information. If investments in databases continue to be destroyed, there will be fewer and fewer people willing to make the investment necessary to create and disseminate these valuable database. Moreover, those database compilers who decide to stay in the database business will be reluctant to expose their most valuable wares to a thieves' market on the Internet. They will keep tight technical or contractual security on their products and only make them available to smaller, more lucrative markets. Inevitably, this will result in fewer and less reliable databases and thus, less information to fuel the information age.

Despite the acknowledged value provided by America's databases, there is presently a lack of meaningful national legal protection for these databases. While database producers rely on several potential legal theories, none adequately nor effectively deter or prevent database

piracy. Depending on the facts of a particular case, database producers may consider raising claims under: U.S. copyright law, state misappropriation law, state trespass law, state contract law, the Computer Fraud and Abuse Act, federal or state trademark or unfair competition law, and trade secret law. However, none of these claims are sufficient. A list of possible claims and an explanation of their shortcomings is provided in the attached appendix, including references to examples where appropriate. The cases identified here are not an exhaustive list of cases illustrating the shortcomings of the current environment, nor do they capture the scores of instances in which the lack of adequate remedies has prevented database producers from bringing suit to combat known instances of piracy.

In addition to the database piracy cases that have resulted in litigation, there are numerous other instances of piracy that never make it to the court room. Many database producers are unwilling to spend significant amounts of money litigating the questionable causes of action discussed above with the very real possibility that they will lose the case and, in the process, draw attention to the vulnerabilities of the company's databases. For example, National Ad Search (NAS), a Wisconsin-based company, takes and uploads employment classifieds from print newspapers in the top 60 markets and sells them to job seekers. It has no authorization to do so (by contract or otherwise). This type of piracy results in loss of good will of advertisers and customers. Newspapers get complaints from advertisers who place classified ads who continue to get harassed by phone calls long after the ad was published and the job has been filled. Cease and desist letters have failed to deter this company which continues to operate today -- NAS refuses to recognize the newspapers' copyright protection and the newspapers have not found a viable state to bring a misappropriation action. Furthermore, neither the Computer

Fraud and Abuse Act nor section 1201of the Copyright Act would apply because the initial takings are both non-electronic and unencrypted.

Based on the long list of database piracy cases that have occurred over the course of the past ten years or so there can be no doubt that there is a definite and significant need for database protection legislation. Moreover, the risk of potential future instances of database piracy and the adverse affects such piracy would have on investments in databases and consumer protection is certainly sufficient justification for Congress to enact database protection legislation. Congress has a long history of legislating to protect against potential future harms. In fact, just last session, Congress passed the TEACH Act creating a new exemption in the copyright law in order to encourage certain educational institutions to create distance education courses. There was no demonstrable evidence establishing that an exemption was necessary, but Congress chose to enact the law based on anticipated – rather than actual – problems using copyrighted works for these courses. There is no reason to think that database protection legislation should be treated any differently.

To be clear, we are not seeking "copyright plus," to expand copyright law, to acquire exclusive rights in the database or to lock up information. We are merely trying to protect against free-riders taking our databases and making them available in a way that hurts our businesses. We think this is a reasonable request. If legislation passes that provides a meaningful legal foundation to fall back on when someone steals a database, companies will be more willing to provide widespread access to their databases and take the risk that it might be stolen. Without this legal basis they are forced to be more cautious about how they disseminate

their databases, how much they should invest in maintaining their databases and, in many instances, whether to create a new database in the first case.

We commend Chairman Sensenbrenner and Chairman Tauzin for their leadership in drafting legislation that seeks to address the problem of theft of our nation's databases. The discussion draft, the "Database and Collections of Information Misappropriation Act" -- reflects years of discussions and negotiations between the Judiciary and Energy & Commerce Committee staff and the stakeholders. The legislation they have developed takes a very targeted and narrow approach to addressing the problem of database piracy.

Unlike prior bills that have addressed the problem of database piracy by providing database producer with exclusive rights to control the use and distribution of a database in any context, the draft legislation developed by the two Committee chairmen, is based on a misappropriation approach that only covers acts of making a database available that cause commercial harm to the database producer. More specifically, the draft legislation creates a narrowly focused prohibition that applies only if ten criteria are met: (1) plaintiff's database must contain a "large number of discrete items;" (2) it must be the result of a "substantial expenditure of financial resources or time;" (3) the defendant must make its database "available in commerce to others;" (4) the amount made available must be at least a "quantitatively substantial part of" the plaintiff's database; (5) the defendant must know that he is not authorized to make the database available; (6) the database is made available "in a time sensitive manner;" (7) the database must serve "as a functional equivalent" of the plaintiff's database; (8) in making the database available the defendant must have caused a loss in revenue to the plaintiff; (9) the

loss in revenue must occur in the *same* market that the plaintiff's database seeks to exploit; and (10) letting this act of misappropriation go unpunished would substantially reduce the incentive of the plaintiff to produce (*i.e.*, create and distribute) the database. These ten criteria – all of which must be satisfied for liability to accrue-- set a very high standard for establishing liability under the draft bill. This standard is even higher when one also considers the exceptions to liability contained in the draft legislation.

We believe that some of the substantive provisions of the draft will provide protection against database piracy while also accounting for the legitimate concerns of database users through narrowly-crafted exceptions and limitations on liability. We also have concerns with the language used in some of the provisions of the draft. Most significantly, we believe the language in some of the provisions – notably the preemption and time sensitivity provisions, among others -- is somewhat ambiguous and could cause inadvertent consequences. Equally as disconcerting, is that the discussion draft does not recognize database thefts that cause noncompetitive harms that adversely affect ISPs and others that have commercial databases. We look forward to working with the Committees to ensure that any preemption of state law is narrowly tailored and does not impede effective licensing of databases or other measures that might otherwise be available, and seeking some appropriate clarifications of the text, including protection against database theft when carried out on behalf of parties other than direct competitors.

We know that a few groups, many of whom were part of the process initiated by the two Committees to come up with a compromise text, have voiced their opposition to the discussion draft. In fact, the discussion draft has a very limited and targeted reach. It protects the database

itself, not the information or facts in a database. The focus of the discussion draft is to protect against unauthorized distributions of a database that cause commercial harm, not to prohibit use or extraction of information from a database. The approach of the discussion draft – relying on a standard of "misappropriation" – is precisely the standard that was recommended by many of those writing now expressing concern.[2] Their continued opposition amply demonstrates that they simply do not accept the conclusion that the Chairmen have both reached: that Congress should legislate to improve the legal protection available for databases

Just as importantly, several of the letters sent opposing the discussion draft recognize that many of their concerns have been addressed. In particular, the draft will not affect the day-to-day activities of librarians, researchers, scientists, and educators or impede their ability to obtain and use facts. The legal standards to establish liability are extremely high. This is so that only database pirates fall within the draft's reach. The customary activities of a librarian, scientist, or educator would not fall within the reach of this draft legislation. As an additional safety valve, there is an additional exception in the legislation that ensures that nonprofit librarians, scientists, or educators are not swept up by the general prohibition in the discussion draft.

Throughout the process initiated by Chairmen Sensenbrenner and Tauzin we have stated our intention to get narrowly targeted and meaningful legislation that addresses the problem of database piracy while also addressing the legitimate concerns of the database user community.

[2] In a paper submitted to the two Committees in Spring 2001 groups representing various libraries and universities and industry stakeholders that participated in the database protection discussions and negotiations initiated by the two Committee Chairmen in March 2001 acknowledged that they "would support a true misappropriation bill for example, one which closely follows the historic standards laid out in *NBA v. Motorola* which would be constitutional and would not stifle innovation, would not impede scientific progress, and would not ultimately hurt the growth of exciting new database products."

To the extent that the opponents believe that the draft falls short of this goal, we continue to stand ready to address those concerns in exchange for their support for this important piece of legislation and their recognition that the bill must recognize database thefts that result in noncompetitive harms and address other concerns of the database publishing community on several of the provisions, including preemption of state laws.

We look forward to working with the Congress and the other stakeholders to achieve a legislative solution that eliminates the unfairness we discussed today. Thank you again for all your work on these very complex issues that have arisen before the two Subcommittees and thank you for your commitment and work to address our concerns in this area. I will be happy to answer any questions.

**APPENDIX**

**THE SHORTCOMINGS OF EXISTING LAW**

(1) Copyright Law: Copyright law does not provide adequate protection for most databases. Copyright law does not prevent a person from taking the non-copyrightable contents of a database, repackaging or reformatting those contents, and distributing the "new" database. As set forth in the Supreme Court decision of *Feist Publications v. Rural Telephone Service Co.*, 499 U.S. 340 (1991), copyright law only protects a compilation (*i.e.*, a database) if there is sufficient creativity in the selection, arrangement or coordination of the compilation. Most of the characteristics that make a database valuable and user-friendly—its comprehensiveness and its logical order (whether alphabetical in print products or random in electronic products)—are routinely deemed to involve no "creative" selection, arrangement or coordination by the courts.[3]

[3] Prime examples of why copyright law does not protect most databases can be found in:

*Warren Publishing, Inc. v. Microdos Data Corp.*, 115 F.3d 1509 (11th Cir. 1997) *vacating* 67 F.3d 276 (11th Cir. 1995). In this case the Eleventh Circuit reversed the district court's order granting an injunction against defendant Microdos Data for copyright infringement of plaintiff's compilation - a directory of information on U.S. cable television systems. Plaintiff's database contained extensive information on cable systems. The entries were arranged state by state in alphabetical order, and within each state, all of the communities receiving cable television services were also listed alphabetically. Plaintiff collected and arranged all of the information about cable systems by selecting the principal community served by a particular cable operator and then cross-referencing additional communities to the principal community listing. Defendant copied plaintiff's directory into its software package. That software package allowed users to rearrange the data in a format of their choosing and to construct searches of the database.

The Eleventh Circuit found that the plaintiff's selection of cable systems contained in its database did not meet the originality requirement under copyright law because the database included all cable systems and, thus, there was no creativity in the selection of what systems were included in the database. In effect, the plaintiff could not prevail because its database was comprehensive. Since the defendant's software left the arrangement of the system up to the user, even if the court had held the arrangement of the data to be copyrightable (which was unlikely because the data was arranged in alphabetical order), there could not have been a copyright infringement because the defendant did not copy the arrangement.

*Mid America Title Co. v. Kirk*, 59 F.3d 719 (7th Cir. 1995). Mid America, a company that produces compilations of land title data, sued James Kirk for copyright infringement after it discovered that Kirk had copied

For example, (i) if a database includes all the facts on a given topic, the court will hold that there is no creativity in the selection because every item in that universe was selected; (ii) if a database is arranged in an order that is logical and useful to the user, the court will hold that there is no creativity in the arrangement because the order is typical; and (iii) if an electronic database is in random order and arranged by the user according to parameters established by the user, the court will hold that there is no creativity in the arrangement because there is no arrangement at all. Therefore, the more useful, complete, and up-to-date a database is (*i.e.*, the qualities that benefit database users the most), the less likely it is to be protected by copyright.

Even when courts find that a database contains elements of selection, arrangement or coordination that are creative enough to warrant copyright protection, the scope of protection afforded has been extremely narrow. For instance, the usual standard for determining copyright infringement is whether there is "substantial similarity" between the allegedly infringing work and the copyrighted work. However, where databases are involved, the standard is heightened to a "virtually identical" standard. That standard has led many courts to hold that a copyrighted

---

one of its compilations. The district court held that neither the selection nor the arrangement of the Mid America database was sufficiently creative to qualify for copyright protection and thus, there was no copyright infringement. The appellate court agreed, expressly acknowledging that the large "amounts of time and effort which Mid America invested in order to gather and report such information [was] irrelevant."

*Skinder-Strauss Assocs. v. Massachusetts Continuing Legal Edu., Inc.*, 914 F. Supp. 665 (D. Mass. 1995). Skinder-Strauss Associates, the publisher of the Massachusetts Lawyer Diary and Manual, known as the "Red Book," sued its competitor MCLE, which publishes the Massachusetts Legal Directory, known as the "Blue Book," for copyright infringement and unfair and deceptive trade practices under Massachusetts law. In creating its directory, the plaintiff listed actively practicing Massachusetts attorneys and judges alphabetically and by jurisdiction. The court found that such arrangement by geography and selection by active bar membership was "typical for any attempt to compile a legal directory for a certain service area." Consequently, the court found that the plaintiff "did not even exercise a minimal degree of creativity…." and that the directory lacked the requisite originality to be protected by copyright. Moreover, the court held that even though "most of the listings published in the Red Book also appear in the Blue Book" (including the fictitious names (i.e., seeds) planted by the plaintiff), the plaintiff's database was not copyrightable; and thus, there could be no copyright violation. In addition, the court held that plaintiff's claims for unfair and deceptive trade practices were preempted by copyright law.

database is not infringed even when the differences between the original database and the copied database are trivial. Courts have also ruled that no infringement occurs when any elements of selection, coordination, or arrangement of the database that was copied did not constitute creative authorship. Accordingly, while some databases may receive copyright protection in theory, in practice the scope of this protection has proven to be minimal.[4]

It has been suggested that Section 1201 of the U.S. Copyright Act would provide a sufficient remedy against database piracy. There is no legal or factual support for this conclusion. Section 1201 was enacted as part of the Digital Millennium Copyright Act (DMCA). It prohibits both the act of circumventing technological protection measures to gain unauthorized access to copyrighted works and the trafficking in any anticircumvention tools that permit unauthorized access. This provision does not come remotely close to addressing the real problem of database piracy because, most significantly, it only applies when the underlying work that is protected by the technological safeguard is a copyrighted work. As noted above, many

---

[4] One example of this occurred in *EPM Communications Inc. v. Notara Inc.*, 56 U.S.P.Q.2d 1144 (S.D.N.Y. 2000), where the U.S. District Court for the Southern District of New York found insufficient evidence of similarity to issue a preliminary injunction against a website accused of infringing a printed compilation of licensing sources. EPM Communications Inc. sued Notara Inc. for copyright infringement, alleging that Notara copied substantial portions of its "Licensing Letter Sourcebooks" in Notara's www.notara.net website. Citing the Feist decision, the court found at the outset that EPM's Sourcebooks were entitled to copyright protection as factual compilations because their creation required the selection of businesses of interest to persons involved in licensing. That selection and subsequent arrangement, when viewed in the aggregate, was sufficiently original to be copyrightable.

As to infringement, however, the court denied the plaintiff's request for a preliminary injunction even after acknowledging that Notara copied fifty-five percent (55%) of the listings in EPM's Sourcebook. The court explained that comparison of a printed compilation and an electronic data base was difficult because the arrangement of the printed compilation could not be perceived in the electronic database unless someone uses the computer "to re -arrange the [database] material into the [Sourcebook] copyright holder's arrangement."

Another recent example of this occurred in *Schoolhouse Inc. v. Anderson*, 2002 U.S. App. LEXIS 63 (8th Cir. 2002). Schoolhouse publishes a magazine marketed to prospective home buyers. The magazine features tables of information on area public and private schools. On January 9th, 2002, the U.S. Court of Appeals for the Eighth Circuit held that defendant's copying and posting of approximately 74%-87% of plaintiff's database of school information on its website did not infringe the plaintiff's copyright in the selection and arrangement of its database.

databases are not protected by copyright and, therefore, could not receive the protections afforded by Section 1201. In sum, the same problems that exist with regard to protecting databases under copyright apply to database providers seeking to assert claims under Section 1201. In fact, those difficulties are elevated in a Section 1201 claim because not only does the database producer have to successfully leap the hurdle of proving that the database was protected by copyright, but also must prove that a technical measure was circumvented in violation of the criteria set forth in Section 1201. In addition, section 1201 provides no remedy against a person who distributes a pirated database that was received from a person who circumvented a technical measure to get it.

Furthermore, access control measures protected under the DMCA, like other technology-based solutions, are, at best, only a partial solution. Technical measures do not work at all where the database is in nonelectronic form, such as classified ads in print newspapers, or directories, such as the McGraw-Hill Companies' World Aviation directory. Similarly, it would not apply to Internet companies, such as eBay, Reed-Elsevier and ExpertPages.com, that elect to allow their customers and users to have open access to some or all of their databases. So, while technological protections may be useful in some business models, in many others they are not.

(2) State Misappropriation Law: State misappropriation law does not provide meaningful national protection to databases. First, each state's misappropriation law -- which is usually a common law doctrine -- is different. In the Internet environment this proves problematic. For example, when a company makes its database available over the Internet,

---

The court found that, although the defendant admitted to copying the database, much of what was copied was attributable to an inevitable selection and obvious arrangement of topics that lacked copyrightable originality.

should it prevent persons from states without misappropriation laws or with inadequate misappropriation laws from accessing its database? Even if it wanted to prevent such access, how would it do so? National uniformity is clearly needed in this area.

Second, many state misappropriation laws are restricted to "hot news". Under these regimes, a database might be protected if it contains "hot news," but only for a short period of time, such as the first fifteen minutes after its inception. The great majority of databases, however, have a value long after the fifteen minutes have expired.

Third, because state misappropriation laws are largely common law, many courts have held their respective state's misappropriation laws to be preempted by federal copyright law.[5] Thus, there must be federal legislation that addresses the misappropriation of databases.

(3) State Trespass Law: State trespass law provides a remedy against database piracy only in the rarest of cases.[6] A significant limitation on state trespass claims is that they do not

---

[5] Another example of database piracy for which a misappropriation claim (and several other state claims) proved unavailing occurred in *Information Handling Services, Inc. v. LRP Publications*, 54 U.S.P.Q.2d 1571 (E.D. Pa. April 18, 2000), 2000 Copr. L. Dec. P28,177 (E.D. Pa. Sept. 20, 2000). Information Handling Services (IHS), a company that creates value-added databases of publicly available government information, brought the action when it found that LRP had copied its product and produced a less expensive alternative. Consequently, IHS "suffered significant losses." As a result, IHS sued on several theories, including violations of state unfair competition laws, misappropriation laws, and trade secret laws. However, the District Court for the Eastern District of Pennsylvania held that each of these state law claims, except for the contract claims, were preempted by the Copyright Act. In addition, the court expressed doubt that the contract was enforceable, but because the court was merely deciding a motion to dismiss, it determined that the ultimate enforceability of the contract did not need to be decided. See also *Ticketmaster v. Tickets.com*, discussed below, and *Lowry's Reports, Inc. v. Legg Mason, Inc.* 186 F. Supp. 2d 592 (D. Md. 2002), in which Legg Mason employees posted a single subscription email report to a company intranet and distributed the report internally, a federal court in Maryland held that Lowry's hot news misappropriation claim was preempted by copyright law. The court stated that "'[f]ree-riding'…, the only element that constitutes a wrongful act, seems indistinguishable from the right to reproduce, perform, distribute or display a work. The other elements do not describe any behavior at all. The cost of generating the information, its time sensitivity, and direct competition between the parties merely define pre-existing conditions; the threat to the plaintiff's business merely 'identifies a consequence of the act of 'free-riding'."

apply to non-electronic databases or databases distributed on CD-ROM. Also, trespass claims will likely only be successful where a plaintiff can prove server or network damage. Most database publishers are not likely to be able to prove this.[7]

In addition, because of the novelty of applying state trespass claims to the Internet, there is no guarantee that other states will interpret these claims the same way that the district courts in California and New York did. As in other cases of reliance on state law, there are substantial variations among states, and national uniformity is needed. Once again, this highlights the need for a uniform federal law providing meaningful database protection.

---

[6] See *EBay v. Bidder's Edge*, 100 F. Supp. 2d 1058 (N.D. Ca. 2000). Unlike real-time aggregators, which merely search the relevant web site pursuant to a search request, Bidder's Edge (BE) copied eBay's entire databases by employing automated data extraction tools, called "web crawlers," "robots," or "spiders" to extract eBay auction listings for posting on BE's site. eBay brought several claims against BE, including a trespass claim under California law. The court agreed with eBay's trespass claim, holding that, if BE's crawling activities were allowed to continue unchecked, it would encourage other auction aggregators to engage in similar searching activities. The cumulative effect caused by similar auction aggregators could be severe, possibly causing eBay's system to suffer gaps in service crash altogether (despite the fact that BE's present activities were found to have a minimally detrimental effect on eBay's site). As evidenced in other court cases since this decision, the eBay decision has no applicability to most of the databases on the market. It applies only to those companies that find their services "crawled" by electronic agents to such a magnitude that the hosting system or network may become overburdened. It also important to note that the injunction issued in this case: (1) does not prevent the pirate from distributing the information it extracted, (2) does not apply outside the state of California, and (3) does not protect database publishers who distribute their databases on printed materials, CD-ROMs, or other traditional media. See also, *Register.com v. Verio,* 126 F.Supp.2d 238 (S.D.N.Y. 2000), on appeal.

[7] The limitations on a trespass claim can be seen in the case of *Ticketmaster v. Tickets.com*, 2000 WL 1887522 (C.D. Cal. August 10, 2000); 2001 WL 51509 (9th Cir. Jan 8, 2001); 2003 Cop. L. Dec. P. 28,607 (C.D. Cal. March 7, 2003). The facts are as follows: Tickets.com established a deep link to the concert information located on Ticketmaster's web site. When Ticketmaster became aware of the deep link, Ticketmaster sued Tickets.com under various claims based on the unauthorized deep linking. After the court ruled against Ticketmaster, Ticketmaster implemented technology which prevented deep-linking to its concert listings. Tickets.com thereafter found a way to circumvent this technology by using "spiders" or web crawlers to copy Ticketmaster's internal web pages, extract the concert information (such as the date, price, time, venue, and band playing) and post it on Tickets.com's web site in a different format. Ticketmaster then sued Tickets.com for copyright infringement, breach of contract, misappropriation and trespass. None of these claims were successful and no injunction was issued. With regard to the trespass claim, the court found that the facts presented by Ticketmaster were compellingly different from those offered by eBay in its lawsuit against Bidder's Edge. Unlike the situation in the Bidder's Edge case, Ticketmaster could not "present the specture [sic] of dozens or more parasites joining the fray." Ticketmaster subsequently lost its



Mr. SMITH. Mr. Wulf.

## STATEMENT OF WILLIAM WULF, PRESIDENT, NATIONAL ACADEMY OF ENGINEERING AND VICE CHAIRMAN, NATIONAL RESEARCH COUNCIL

Mr. WULF. Thank you, Mr. Chairman. Like my predecessors here, I am very grateful to have the opportunity to testify on this important legislation today. I should clarify that I am testifying on behalf of the U.S. National Academies. That is The Academy of Sciences, The Academy of Engineering, and institute of Medicine, but also on Behalf of the Association of American Universities, the

American Library Association and the Association of Research Libraries.

The proposed legislation concerns a topic about which the scientific research, education and library communities have had an abiding and continuing concern. For all of those communities, the free and open sharing of information is essential for progress, and that progress is in large measure responsible for our quality—you are improving quality of life. It is estimated that one half of our GDP growth is due to advances in science and technology. It is estimated that two-thirds of our productivity growth, the real source of wealth creation, is due to advances in science and technology.

A hundred years ago this year, the Wright brothers flew the first heavier-than-air aircraft. A hundred years ago this year, Henry Ford opened his mass production plant making affordable automobiles available to all of us. A hundred years ago this year, the average life expectancy of an American was 46. It is now 77 an increase of 31 years.

On the light side, I suppose it is interesting to speculate how many of us in this room would be here if it weren't for those advances in science and technology.

On the serious side, my point is that this is a system that one tampers with very carefully. An unintended consequence of well-intended legislation on the free and open sharing of information could have enormous implications on our prosperity, our health and indeed on our national security.

I would just like to make a few points about the draft legislation. First of all, I want to acknowledge that, in fact, the draft has a number of significant improvements, and we are very grateful for that. I must admit that my own analysis is incomplete, but I believe that there are also still problems and ambiguities, some of which are in my written testimony. As a consequence, the Academy has remained committed to being helpful in producing a balanced and fair bill, assuming that one is deemed necessary.

The key I think is that there are several principles that ought to inform the process of crafting any new legislation in this area, and I would like to focus the rest of my testimony on those principles.

The first is that the public domain status of factual noncopyrightable information must be preserved. Any new protection regime should leave a wide buffer zone to ensure that factual information will not be subjected to proprietary claims.

Two, only significant problems of unfair competition in market failure that have been proven should be addressed. And negative unintended consequences must be avoided.

Three, a reasonable balance of interest among the stakeholders in an information economy should be maintained. Congress should proceed cautiously in creating new protection regimes, because once created, a new protection regime is virtually impossible to dismantle.

Four, healthy competition in the information industry needs to be promoted, while the further strengthening of unwarranted monopolies should be avoided.

Five, exclusive control by private parties over information in databases produced by the Government must be prevented.

Six, new protection regimes should not create any doubt or controversy about the lawfulness of traditional and customary access to and use of factual information for not-for-profit science, research and education, effective exceptions must be adopted.

And finely, the important role and function of our Nation's libraries must not be undermined.

Before concluding, let me note that there is little evidence, as far as I can tell, since the last time we testified on this issue that databases or other collections of information are routinely stolen or that there is a significant market failure in the information industry for such products.

Indeed, database producers have already enjoy a broad range of legal technological and self-help methods, many of which have been further strengthened in recent years, that protect the fruits of their investment.

Nevertheless, the academies and the other organizations represented in this testimony remain committed to playing a constructive role in helping Congress to consider the issues of database protection in a way that is consistent with the principles I have identified above. Thank you.

Mr. SMITH. Thank you, Mr. Wulf. And thank you all for your very strong testimony today.

[The prepared statement of Mr. Wulf follows:]

PREPARED STATEMENT OF WM. A. WULF

INTRODUCTION

My name is Bill Wulf. I have been asked to testify on behalf of the U.S. National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine (the "Academies"). As you know, the three Academies were chartered by Congress to provide advice to the federal government and to the nation on scientific, technical, and medical issues. My testimony is also being given on behalf of the Association of American Universities, the American Library Association, and the Association of Research Libraries.

I am grateful to have the opportunity to testify to you today about the draft legislation called the "Database and Collections of Information Misappropriation Act of 2003." This proposed legislation concerns a topic about which the scientific, research, education, and library communities have had an abiding interest and continuing concerns. Indeed, this is the third time that the Academies have testified on congressional legislation in this area since 1997, and both the Academies and their operating arm, the National Research Council (NRC), have published extensively on these issues over the past seven years. A list of recent relevant NRC reports and my biographical summary are provided at the end of this statement. Copies of the referenced NRC reports, as well as the Academies' previous testimony, letters to Congress, and background analyses that we have written on previous versions of this legislation, are available on request.

Although I am authorized to speak only on behalf of the organizations that I represent here today, the issues I wish to raise with you pertain broadly to our nation's scientific, research, education, and library concerns. And although I do not address directly the important issues raised by this legislation for the commercial sector, which are the focus of other testimony before you, my remarks are cognizant of the broader implications to our nation's economic and social progress.

My testimony makes the following points, which build on our previous analyses:

► As a matter of public policy, there are several key principles that must inform the process of crafting any new legislation in this area, including the following:

1) The public-domain status of factual, non-copyrightable information must be preserved, and any new protection regime should leave a wide buffer zone to ensure that factual information will not be subjected to proprietary claims.

2) Only significant problems of unfair competition and market failure that have been proven should be addressed, and negative unintended consequences must be avoided.
3) A reasonable balance of interests among all stakeholders in the information economy should be maintained. Congress should proceed cautiously in creating new protection regimes, because once created, a new protection regime is virtually impossible to dismantle.
4) Healthy competition in the information industry needs to be promoted, while the further strengthening of unwarranted monopolies should be avoided.
5) Exclusive control, either de jure or de facto, by private parties over information and databases produced by the government must be prevented.
6) New protection regimes should not create any doubt or controversy about the lawfulness of traditional and customary access to and use of factual information for not-for-profit science, research, and education. Effective exceptions must be adopted.
7) The important role and functions of our nation's libraries must not be undermined.

► The draft legislation includes a number of improvements over previous versions of this legislation that have been introduced by the House Committee on the Judiciary since 1996.

► There are still major problems and ambiguities in the current draft bill that can and should be addressed, assuming that the creation of a new statutory remedy is still deemed necessary.

► The Academies and the other organizations represented in this testimony remain committed to playing a constructive role in helping Congress to consider the issues of database protection in a way that is consistent with the principles identified in this testimony and that avoids negative unintended consequences.

* * *

A. KEY PRINCIPLES

1) The public-domain status of factual, non-copyrightable information must be preserved, and any new protection regime should leave a wide buffer zone to ensure that factual information will not be subjected to proprietary claims.

As we have noted in previous testimony on this issue, access to and use of factual data in the public domain is essential to furthering our understanding of nature, to the validation of scientific claims, and to the progress of science and our nation's system of innovation. The 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.

Not only researchers and educators, but all citizens with access to computers and networks, constantly create new databases and information products for both commercial and noncommercial applications by extracting and recombining public-domain data and information from multiple sources. The rapid and continuous synthesis of disparate data by all segments of our society is one of the defining characteristics of the information age. Moreover, the serendipitous nature of research and the need of scientists and others to make transformative uses of non-copyrightable facts are such that one cannot predict when or how a database will be used. The ability of individuals and organizations to use information in a wide variety of innovative ways is also a measure of success of the original data-collection efforts.

Society uses the fruits of such research and innovation to expand the world's base of knowledge and applies that knowledge in myriad downstream applications to create new wealth and to enhance the public welfare. Indeed, the policy of the United States has been to support a vibrant research enterprise and to assure that its productivity is exploited for national gain. Thus, freedom of inquiry, the availability of scientific and other factual data in the public domain, and the open publication of

results are cornerstones of our research system that U.S. law and tradition have long upheld.

The results of these wise policies have been spectacular. For many decades, the United States has been the leader in the collection and dissemination of scientific and technical data and in the discovery and creation of new knowledge. Our nation has used that knowledge more effectively than any other nation to support new industries and applications, such as the biotechnology industry and the discovery of new diagnostics and cures for hereditary and other diseases.

In addition to the critical importance to our progress in science and innovation for factual information to remain in the public domain, it also is essential for many other compelling American values and needs, including 1st Amendment rights of freedom of expression, the promotion of the information economy, democracy and good governance, and other public- interest uses by consumers and society generally.

Because of the overriding importance of non-copyrightable factual information remaining in the public domain, any new legislation in this area must be limited to remedying unfair conduct in commerce rather than extending any exclusive property rights in the factual information itself.

Where there is uncertainty or doubt about the effect of potential new legislation, Congress should be careful to err on the side of caution. When the subject matter consists of the fundamental building blocks of knowledge, science and expression, the cost of over-protection far exceeds the cost of under-protection.

2) Only significant problems of unfair competition and market failure that have been proven should be addressed, and negative unintended consequences must be avoided.

Proponents of new database protection legislation have long argued that the misappropriation of databases is a major problem in the U.S. information industry and that existing methods of protection and remedies are inadequate. We find both assertions to be of increasingly dubious validity.

There is little evidence since the last time we testified on this issue before Congress that databases or other collections of information are routinely stolen or that there is massive market failure in the information industry. Indeed, database producers already enjoy a broad range of legal, technological, and self-help methods-many of which have been further strengthened in recent years-that protect the fruits of their investments. Available legal remedies at the federal level include traditional copyright law, new rights to prevent the circumvention of technological protection measures granted under the Digital Millennium Copyright Act, and the new Computer Fraud and Abuse Act. Under state law, many jurisdictions have a common law prohibition against misappropriation of "hot news," and a claim for trespass to chattels to protect databases.

Contracts and licenses are now used universally by database owners to make their products available under a range of custom-tailored, restrictive conditions. Technologies that protect digital databases and help enforce the existing statutory and contractual rights of owners are constantly being refined and strengthened, including such methods as encryption, online database access controls, software and hardware based trusted systems, and digital object identifiers and electronic watermarks. Indeed, these contracts and technologies are increasingly employed to limit uses of data and information that would otherwise be permitted by law. Congress should carefully monitor their use and consider whether limits on their use are needed to preserve the balance between access to and use of factual information and the incentives to invest in the collection of such information, both of which are essential to the vigorous growth of science and knowledge.

Finally, market based protections of databases through self-help business practices such as frequent updating and customizing can help make misappropriation less effective. Taken together, these database protection methods have helped make the commercial database market expand successfully in the United States.

The Academies, the Association of American Universities, the American Library Association, and the Association of Research Libraries nonetheless are committed to playing a constructive role in helping Congress to consider the issues of database protection in a way that is consistent with the principles identified in this testimony and avoids unintended negative consequences. The National Research Council reports referenced at the end of this testimony analyze the far- reaching negative implications to research and innovation that could result from legislation that is overly protective of data and non-copyrightable factual information.

3) A reasonable balance of interests among all stakeholders in the information economy should be maintained. Congress should proceed cautiously

in creating new protection regimes, because once created, a new protection regime is virtually impossible to dismantle.

It is essential to consider fully and to promote a healthy balance of the interests of all the stakeholders in the information economy and society, including the general public. The trend in recent years has been to increase the breadth, depth, and length of all types of intellectual property protection. The creation of any new statutory rights, particularly for subject matter as sensitive as non-copyrightable factual information, must be done in full cognizance of the interaction of these rights with other parallel rights conferred by other statutes to avoid negative synergistic effects. In this regard, a major concern for the research community, as discussed further below, are the potential negative effects on access to and use of databases from unbridled, highly restrictive licensing practices, especially through increasingly legitimized adhesion contracts (e.g., shrink-wrap and click-on licenses), in concert with any additional new statutory rights in databases.

Further, history has demonstrated that once granted, intellectual property rights are rarely, if ever, reduced or limited. Thus, if there is uncertainty about the effect of any proposed new protection, it is important err on the side of caution and the preservation of the status quo.

4) Healthy competition in the information industry needs to be promoted, while the further strengthening of unwarranted monopolies should be avoided.

The promotion of competition is primarily an economic issue of direct interest to our colleagues in industry, but the benefits of competitive prices and increased quality accrue to the public. It is important, nonetheless, to emphasize that a preponderance of scientific databases are produced by sole sources, whether in the public or the private sector. For example, the vast majority of observational data sets of phenomena in the natural world, as well as all unique historical factual compilations, can never be recreated independently and are therefore frequently available only from a single, original source. In other cases, scientific databases are de facto unique natural monopolies because the cost of producing the data and the potential market are such that the economics will not support multiple sources. Even when data that are similar, but not identical, to original research results or observations are available for use in non-technical applications, researchers and educators are unlikely to consider an inexact replica of a database to be a suitable substitute if it does not meet fully the original specifications. For this reason, scientific databases are particularly prone to monopoly control. Any new legislation therefore must not enhance the market power of sole-source providers in any segment of the information industry without adequate public-interest safeguards.

5) Exclusive control, either de jure or de facto, by private parties over information and databases produced by the government must be prevented.

Consistent with principle #1 above, the public domain status of governmental databases and other information products is a key factor for the success of our nation's research enterprise, as well as for other compelling national values and interests. Legislation that confers new rights on the private sector must fully exempt government databases from the scope of protection and avoid the possibility of exclusive capture by private-sector entities.

6) New protection regimes should not create any doubt or controversy about the lawfulness of traditional and customary access to and use of factual information for not-for-profit science, research, and education. Effective exceptions must be adopted.

Also in keeping with principle #1 above, it is important to provide clear immunity for customary non-commercial scientific, research, and educational uses from the scope of a database protection statute. Non-profit institutions should not be required to have expert intellectual property counsel looking over the shoulder of every scientist and scholar. Customary activities should not be chilled. Because in the case of databases, facts themselves are at issue, the legislation should include an express presumption that such customary uses are exempt from liability and the burden of proof on the plaintiff of demonstrating a violation should be heightened.

7) The important role and functions of our nation's libraries must not be undermined.

Libraries traditionally have served the important public-interest function of providing access to information to our nation's citizens, and performed essential preservation and archiving activities. Any new rights conferred by new legislation on data-

base owners must not undermine the libraries' ability to continue its role as public-interest intermediary for the access to and preservation of factual information resources.

* * *

B. PRELIMINARY COMMENTS ON THE DRAFT LEGISLATION

We have not had sufficient opportunity to analyze comprehensively the draft "Database and Collections of Information Misappropriation Act of 2003." The issues and competing interests in this legislation are complex and difficult to reconcile. Although the process has been long and difficult, we believe that it has led to a deeper understanding of the issues, which was so palpably lacking when the first legislative proposal, based on the European Union's database directive, was introduced in 1996. It also has demonstrated the inherent problems with introducing any new rights in this Constitutionally sensitive area and the importance of addressing adequately the competing legitimate interests of the many stakeholders in the information economy, not only the economic interests of the originators of commercial databases.

Our preliminary analysis of this new version of the legislation is consistent with the views expressed by the major university organizations in the September 9, 2003 letter from Nils Hasselmo, President of the Association of American Universities, to the two cognizant Committee Chairmen. We conclude that although improvements have been made over the previous legislative proposals introduced by the Committee on the Judiciary, very significant problems still remain to be resolved. Moreover, the current draft contains a number of new provisions whose intent and impact are ambiguous and which could have serious unintended consequences for the research and education enterprise.

We appreciate, in particular, several improvements that have been made in response to the concerns expressed earlier by the Academies and other parties to this process. The move toward a standard of liability grounded more in unfair competition law and the elimination of some of the most unacceptable aspects of previous versions of the Committee on the Judiciary's proposed statutes, are certainly welcome. Among the specific improvements that we see are the elimination of qualitative substantiality, the effort to tie liability to direct competition in the same market as the existing database, the adoption of a knowledge requirement as a condition of liability, and a limitation to databases that require substantial effort to develop. The elimination of criminal penalties and the explicit recognition of the doctrine of misuse as a limiting factor on lawsuits are also positive developments.

Although the discussion draft addresses some of the concerns we identified previously, many serious problems remain nonetheless, while new ambiguities have been introduced by the recent changes. We note here only the issues of greatest concern to the scientific, research, education, and library communities, consistent with the principles articulated above, and also incorporate by reference the additional concerns expressed in the September 9 letter from Nils Hasselmo. In particular:

- With regard to the liability standard, the discussion draft could confer perpetual ownership rights in a wide variety of data by virtue of protecting investment based on open-ended maintenance of a database. In addition, the concept of "making available to others" appears to be overly broad, posing a threat to customary collaborative work within or among universities and research institutions. Moreover, a minimal amount of harm-even one lost sale or a single lost source of data-could lead to a finding of liability and to a chilling of the use of public-domain factual information, contrary to the values articulated under principle #1 above.
- The exception for educational, scientific, and research institutions applies only if the institutions are nonprofit and their "making available" is for nonprofit purposes. This would discourage joint research and development activities between nonprofit institutions and corporations. Especially troubling is that the exception can be overridden by a shrink- wrap or click-on license and render the exception meaningless—a major concern noted under principles #3 and 6. Any new legislation must preclude such a possibility. Finally, we continue to urge that the burden of proof of demonstrating that customary not-for- profit scientific, research, and educational uses of factual information are unreasonable should be a heavy one and should be borne by the plaintiff.
- The scope of the exclusion for government information in the discussion draft is uncertain as well. It appears that a publisher that incorporates gov-

ernment information in its database could prevent others from making available that government information - even if it is not available from any other source, contrary to principle #5.

► By failing to address the problem of sole-source databases, the discussion draft increases monopolists' control over competitive uses of information. This is of particular concern in the market for databases used in scientific research and education, as noted under principle #4. The provision on misuse, which could help mitigate harmful conduct of database monopolists, lacks any guidance for courts to determine whether misuse occurred. The misuse provision should specifically address the issue of sole-source databases. H.R. 1858 contained appropriate language in this regard.

While we believe that the Committees have made progress on this legislation, it is clear that the current discussion draft is still not ready to be adopted and would introduce serious problems in its present form for many stakeholders in the information economy, including the scientific, research, educational, and library sectors.

In closing, I would like to reiterate that the Academies, and all of the organizations I represent in my testimony today, have sought to play a constructive role in the congressional efforts to craft appropriate legislation in this complex and sensitive area. We look forward to working with Congress on this issue to develop a consensus on how best to move forward from here.

Thank you again for providing us with the opportunity to testify at this hearing.

* * *

Recent relevant National Research Council reports, published by the National Academies Press and all freely available at: www.nap.edu :

*The Role of Scientific and Technical Data and Information in the Public Domain (2003)*

*The Digital Dilemma: Intellectual Property in the Information Age (2000)*

*A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases (1999)*

*Bits of Power: Issues in Global Access to Scientific Data (1997)*

Mr. SMITH. At this point, I am going to turn the Chair over to Congressman Stearns, and we will continue with the hearing.

Mr. STEARNS. I thank my colleague, and I certainly would defer to him to start with his questions.

Mr. SMITH. Thank you, Mr. Chairman. Mr. Carson, let me direct my first question to you, and this goes to page 8 of Mr. Wulf's prepared testimony, and he mentions several examples of serious problems with the proposed legislation, and let me read you the first two and ask you to tell us why they are or are not serious problems.

Now, he says, first with regard to the liability standard the discussion draft could confer perpetual ownership rights in a wide variety of data by virtue of protecting investment based on open-ended maintenance of a database. And second, the exception for educational, scientific and research institutions arise only if the institutions are nonprofit and their making available is for nonprofit purposes.

Do you consider those to be serious problems, and if not, why not?

Mr. CARSON. We understand the concern, Mr. Chairman. However, we are not certain how serious they are in the context of this bill. Let's take the first one. Would this discussion draft confer perpetual ownership rights? We don't think so, and we don't think so primarily because of one of the requirements that the unauthorized making available in commerce has to occur in a time-sensitive manner.

As we stated in our testimony, there is, we think, some ambiguity in the current draft that suggests what is meant by that. If, as we understand, this is simply a codification of the "hot news" doctrine, then we don't see how there can be any conceivable problem with perpetual ownership rights. If it goes beyond "hot news," then we really have to give it further thought and have a clear understanding of what is meant by this time-sensitive component and the elaboration that a court should consider the temporal value of the information in the database within the context of the industry sector involved. That, we think, does need some amplification and clarification, and depending on where that leaves us, we may or may not have some kind of problem.

On the second aspect, whether the restriction of the exception for educational, scientific and research institutions to nonprofit institutions is a problem, I guess my first reaction to that would be that while we think it is a good thing that this legislation—this proposed legislation—would have an exclusion for nonprofit educational scientific and research institutions, when you look at the scope of the prohibitions set forth in section 3, which is clearly addressed to competitive activity which is serving as a functional equivalent in the same market as the data base—the proprietor's database—it is really hard to imagine too many situations, if any, when a nonprofit, scientific or educational research facility would be engaging in an act which would even implicate that primary prohibition.

So that would be the first part of my answer, and the second part would be—we think there is a distinction between nonprofit institutions and for-profit institutions. If you are in this business for a profit, then maybe you should be on the same playing field as any other for-profit actors.

Mr. SMITH. Thank you, Mr. Carson.

Mr. Kupferschmid, in the appendix to your prepared testimony, you give several examples of databases that are not currently protected by copyright law. Two of the examples you gave are the directory of information on U.S. cable television systems, and the second was a Massachusetts lawyer diary and manual.

My question for you is why should they be protected? And my question for Mr. Donohue and Mr. Wulf is why not?

Mr. KUPFERSCHMID. One of the cases you are talking about, *Warren Publishing* v. *Microdos*, was a case involving a cable directory, and in that case Microdos came along and copied and sold the database in competition with Warren, a typical case of free riding, the exact type of situation we are trying to address here.

Since that case has come down Warren Publishing has taken steps to try to protect themselves. What they have done is they have now put a shrink wrap license around the cable directory, because they feel that is pretty much the only way that they can protect themselves.

I think it is quite interesting that because there is no law to protect themselves and they have taken this step, they have gotten this directory sent back to them from libraries and from others who do not want to adhere to the shrink wrap agreement. They won't open it up. So what that means is because there is no law and because Warren Publishing has had to take these steps to protect

themselves, in essence, there is information that is not getting out there. And that is because there is no law and they have had to use other means to protect themselves. If there were a law, then presumably they would not have to do that and——

Mr. SMITH. Thank you. Mr. Donohue and Mr. Wulf, very briefly, why shouldn't we protect that directory and that—directory of information U.S. cable television systems and the Massachusetts lawyer diary and manual? Why shouldn't we protect those databases?

Mr. Donohue.

Mr. DONOHUE. I believe that there is enough law on the books to give most companies the protection that they would need. What we are doing here, when you think about the Internet, when you think about the databases that are available—and there are millions of them—what we are doing here is talking not about the facts, as my colleague said here, but about the format, how they are put together. And certain of those things ought to be paid for. If they are not paid for appropriately, then the law which we now have should be used to protect those that are injured.

To put together a new piece of legislation, to run around and look within those millions of databases for somebody that has been harmed is—in some ways this is getting ridiculous, because we are looking around for a problem for this solution.

Mr. SMITH. Mr. Donohue, if current law was not sufficient, would you support legislation that would close that loophole?

Mr. DONOHUE. If I saw it and it made sense and if there was a serious enough problem to be resolved, of course.

Mr. SMITH. Okay. Thank you, Mr. Donohue, and thank you, Mr. Chairman.

Mr. STEARNS. I thank my colleague.

Mr. Carson, let me just follow up here a little bit with what my colleague talked about. Now, you expressed the view that misappropriation is the best approach to this issue. Right?

Mr. CARSON. That's correct.

Mr. STEARNS. And so when you say misappropriation, you are talking about products versus rights of facts. Is that a good interpretation of what misappropriation is? In other words, you are saying misappropriation as a rule could be used to say that the facts themselves as developed could be used as property, products. No? Why don't you define misappropriations for me.

Mr. CARSON. No. Misappropriations—in fact on the opposite end of the spectrum from the property right—is an aspect of unfair competition, Mr. Chairman, and therefore the—basically the focus of a misappropriation claim would be the act of competition, the taking of someone else's database and using it in competition with them.

Mr. STEARNS. Okay. Now, in the Supreme Court *International News Service* v. *Associated Press*, do you believe that it is necessary for that approach to closely track the language that is in that decision?

Mr. CARSON. I don't think it is necessary, Mr. Chairman. I think it is probably the minimum. It is my understanding that this discussion draft, in fact, is an attempt and not a bad attempt to track the language in the *INS* decision and subsequently the *NBA* deci-

sion which also follows that, but it doesn't necessarily have to be that narrow.

Mr. STEARNS. And yet in the Supreme Court decision with *Feist*, it says no copyright protection for noncreative databases no property right. Now, does that go against what you just said?

Mr. CARSON. Not at all. It is talking about copyright. We are not talking about copyright here. We are talking about an entirely different species of protection.

Mr. STEARNS. So you would agree that what the bill in tracking the INS decision is the appropriate way to go?

Mr. CARSON. We would think that that is an appropriate way to go, I think, is how I would put it.

Mr. STEARNS. An appropriate way.

Okay. Mr. Donohue, how do you respond to the specific example cited by Mr. Kupferschmid of court cases that have not offered protection to databases? And we hear all the time, you know, that there is no protection for these databases. What would be your response that the courts are not helping out?

Mr. DONOHUE. Well, we have a couple of examples here of legal situations that narrowed the scope of what is protected, and therefore, it is seen to be not helping out. The issue here, you know, we have gone through a phenomena over the last 10 years in this country in the expansion of databases that are no longer held in our hand. We look at them on the screen or we pull them off from an electronic system. We have gone through the whole question of taxing and how do you pay for these issues and the protection and privacy.

This issue is so small compared to everything else, I would not argue with you, Mr. Chairman, that you somewhere can find someone that was injured; but the preponderance of evidence is that the benefits far outweigh the loss in terms of what is happening with this valuable information without intentionally injuring anybody's economic well-being. What is the benefit for our society versus—and you have to ask a question, why do we really want this legislation? Do we want it because every day people have economic loss or intellectual property loss? Or do we want it because certain people are looking for a way to perhaps capture information for their own economic gain? And I am not suggesting that that is the only interest here.

What I am suggesting is that our members, AT&T and Yahoo and Bloomberg and Schwab and people that maintain extraordinary databases are saying, Tom, we are not being injured. We don't really think this legislation is going to help. And what it is going to be is a retirement opportunity for certain class action or mass action or straight-action lawyers. We don't need this type of new legislation, because we don't have a problem to fix.

I wouldn't argue, though, Mr. Chairman, that a very smart guy from a good organization that it itself is divided on this issue can't find an exception or a circumstance where somebody was injured.

Mr. STEARNS. My time has expired. You probably could have touched on the fact that the bill allows for quadruple damages.

Mr. DONOHUE. I said that in my testimony.

Mr. STEARNS. Okay.

Mr. Berman.

Mr. BERMAN. Thank you, Mr. Chairman. Just to clarify, Mr. Donohue, you are not saying that there is something inherently wrong when people try to capture information for their economic gain, are you?

Mr. DONOHUE. No. As a matter of fact, that is what I represent on a fair and equitable way without disadvantaging others through the use of the Government.

Mr. BERMAN. Mr. Carson, the ACLU writes in opposition to the draft legislation they raise first amendment concerns, which I think the Committees should appropriately—should consider and give way to, but they also state that the copyright clause of the Constitution stands as an affirmative bar to protection of unoriginal compilations of facts and that therefore the draft bill violates the copyright clause. While Chairman Stearns didn't quite say that in his opening statement, he came pretty close to saying that. What does the Copyright Office think of this assertion that the copyright clause of the Constitution prevents Congress from protecting facts and that the bill attempts to evade an important Congressional limitation on—constitutional limitation on Congressional power?

Mr. CARSON. Mr. Berman, clearly the copyright clause prevents Congress from protecting facts through the vehicle of the copyright laws, and there is a pretty good argument that the copyright clause might as well prevent Congress from protecting databases through conferring a property right that is essentially the same as a copyright on database producers; but the copyright clause does not speak to Congress's power under the commerce clause to regulate competition in this industry if what the legislation is in fact regulating is that competition and is not conferring exclusive rights, which is the copyright package and which is what the copyright clause, in fact, addresses.

Mr. BERMAN. Thank you.

Dr. Wulf, you argue that new database legislation is unnecessary because the DMCA copyright laws, trespass actions and contracts provide database producers with adequate protection, yet many of the organizations that you represent today advocate in other contexts rolling back the DMCA, weakening copyright law and placing Government limitations on the freedom to contract in the form of compulsory licenses and things like that.

Isn't there an inconsistency between these positions? The organizations you represent appear to be proposing that we take away the very protections that they assert through you, provide adequate protection to database creators. Will they support strong database legislation if they succeed in rolling back protection in these other areas?

Mr. WULF. I think inflating the two issues is probably a mistake. You are absolutely right, there are organizations which have differing views from my own on some of these issues. I happen to have been an entrepreneur who spun out from a university and started a software company and intellectual property rights in software which is very much a gut issue for me.

Mr. BERMAN. I don't think I inflated the two issues. I think the organization you represent has.

Mr. WULF. I am representing those organizations with respect to this particular bill and not with respect to everything that they have ever said or done.

Look, I think the essential point here is—and by the way, you probably know that in earlier testimony we said—we, the academies, not everybody I represent here, said that we thought there was potentially a gap that needs to be filled. Things have happened since then. We need to now, I think, stand back very carefully and ask whether the things that have happened, like the Digital Millennium Copyright Act, have closed that gap to the point where this legislation is no longer needed. I was not advocating not doing it. I am advocating taking a very careful look.

Mr. BERMAN. At whether or not the DMC——

Mr. WULF. Whether that gap still exists.

Mr. BERMAN. And if it doesn't, it is because the Digital Millennium Copyright Act may have closed that gap.

Mr. WULF. And the Computer Fraud and Abuse Act and new technology in cryptography and, you know, a lot of things. It is not the one thing.

Mr. BERMAN. Thank you very much, Mr. Chairman. I yield back.

Mr. STEARNS. I thank the gentleman.

Mr. Keller, my colleague from Florida.

Mr. KELLER. Thank you, Mr. Chairman. Let me begin with Mr. Kupferschmid. I was taking some notes when Mr. Donohue was speaking. He said there is no real world example of a database not protected under current law. I got the gist of his testimony was there is not really a problem, because you have the adequate existing laws to cover it. If you were speaking to a sixth grade class, what would you say is a real word example they could relate to that is not protected by existing law?

Mr. KUPFERSCHMID. Well, the example I would give them would probably be *Ticketmaster* v. *Tickets.com*, because that involved concert information. So why not go with that since we are dealing with a teenage crowd here. In that case, Tickets.com originally deep-linked to the information in Ticketmaster's Web site. Ticketmaster didn't want them to do that, so they were able to use technology to prevent them from deep-linking. The technology didn't work because Tickets.com then came and sent out a Spider, Robot, Web Crawler, whatever you want to call it, and copied the information and put it into their database. So Ticketmaster then sues Tickets.Com since the technology didn't work.

Well, Ticketmaster then sues for four claims. They sue for copyright infringement, misappropriation law, breach of contract and also trespass law. Guess what? They lost on all four claims. They lost on copyright because there was no copyrightable expression in the concert data that was taken. They lost on misappropriation, because there is a "hot news" misappropriation requirement, and the information was not considered to be time-sensitive or highly time-sensitive under that criteria.

Under breach of contract theory, there was held to be no contract. This was just a Web page with terms and conditions at the end. Browse wrap licenses like that are usually not enforceable.

And lastly in trespass, this was the same court that heard the trespass claim in the eBay Bidder's Edge case; but yet, they turned

around and said there was no damage to a server or any type of damage to hardware that Ticketmaster had and therefore there was no trespass. So they tried technology, it didn't work. They tried contract, it didn't work. They tried misappropriation. They tried trespassing. They tried copyright. All out of luck. They lost. That is a typical case.

Mr. KELLER. That is a sophisticated sixth grade class. All right. In fairness to the other side, Mr. Donohue let me ask you a tough question here. You argue that there is essentially not a problem, because existing laws offer protection in a database. At least that is how I inferred your testimony. However, I understand that some opponents, including I guess some of your members such as Bloomberg, are making arguments in court now such as the day star versus 20th century case, the one they signed on to, that the current protections provided by contracts, State laws on trespass and the Computer Fraud and Abuse Act should be precluded from protecting databases.

How do those two positions coexist?

Mr. DONOHUE. A sixth grade answer?

Mr. KELLER. I hope so. That is all I can understand up here.

Mr. DONOHUE. My first argument is there is not a repetitive compelling problem of any size or shape that would compel the Congress of the United States to act and pass new legislation. My second issue is that we heard what the court didn't do and regularly doesn't do with the occasional case that it gets. I think that deserves some consideration in terms of maybe what is happening is people are using, for the most part, information that should be available. I make no argument for even a sixth grader that there aren't occasional problems that we would all feel badly about, but we don't need a massive new set of legislation that is going to give the trial lawyers a whole new retirement program just to deal with those occasional issues. Thank you very much to the sixth grade.

Mr. KELLER. Well, if this legislation had the misappropriations standard, which I think you previously wanted and didn't have, the quadruple penalties and had a couple of other things that were changed, would this be something that you think would merit support or do you think it is just a fatally flawed idea from the getgo?

Mr. DONOHUE. Well, for the reasons I expressed, I rather think it is flawed. Of course, any time the Congress and the Committee with smart people like this change a piece of legislation around and narrow its scope, we would always look at it and we would look at it thoughtfully.

What I am commenting on is the draft that was put before us, and that would be a mistake for this Committee to advance.

Mr. KELLER. Thank you, Mr. Chairman. I have nothing further.

Mr. STEARNS. I thank my colleague.

The Ranking Member on the Commerce Consumer Protection and Trade, we welcome her, Ms. Schakowsky.

Ms. SCHAKOWSKY. Mr. Kupferschmid, Mr. Donohue said that he has members of his association which has taken a very strong position against this legislation who themselves have databases, and I am trying to understand then what the difference is between the interest that you represent and the interest—and I would welcome

Mr. Donohue's comment as well. If one segment doesn't have a problem, are you just finding—you know, as Mr. Donohue has suggested—a couple of examples but that industry wide it is not a major problem that needs our addressing?

Mr. KUPFERSCHMID. I think the difference between the number and the types of database producers that we represent compared to the chamber of commerce, I think it is significantly different. We have in our coalition here 70 companies that are pure database producers that are supportive and they know that there is a need for legislation. I won't speak for the Chamber. I will let Mr. Donohue speak for himself, but I highly doubt that there is that number of companies in his association interested in this issue. I know he referenced the fact that we are divided on this issue. Nothing could be further from the truth. I mean, SIIA and CADP are very supportive of database legislation protection, and recognize the need for it. The coalition against database piracy, which involves members like Dow Jones, eBay, the Newspaper Association of America, McGraw-Hill, the National Association of Realtors, and smaller companies like Carfax and Berkshire Publishing, they are all supportive and there is they believe that there is a definite need. These are companies that their business is producing databases.

Ms. SCHAKOWSKY. Mr. Donohue.

Mr. DONOHUE. Thank you very much for the opportunity to comment. First, we have more than 3 million members. We have a majority of the major database holders as our members. The list that you have been submitted on the other side has 14 subsidiaries of one company, has a number of people, who I think if they had an opportunity to review this legislation might have another view, but let me say, Congresswoman, that the interesting thing about my job is not dealing with the Congress, because that is easy.

It is dealing with my members. But the great majority of our members have made it very clear that an added legislation that opened up this litigation problem and that made it more difficult for people to use their product and access their information would be a difficulty for them. And if the Committee would like, I would be very happy to give you an extensive list. I think it would be a little longer than the one over here.

Ms. SCHAKOWSKY. Thank you. I am looking at the letter from the American Civil Liberties Union, and one of the issues that they raise is the issue of subpoena powers. They say subpoena powers pose serious privacy concerns. Section 7 of the bill gives database owners broad subpoena powers with no judicial discretion. A clerk must grant a subpoena as long as the proposed subpoena "is in proper form," and that "the accompanying declaration is properly executed." there is no prescription for due process protections or even a requirement that an applicant may prime fascia demonstrate that a proposed defendant has, indeed, violated the law before violating her privacy.

This lack of privacy related safeguards seem ripe for abuse. I don't know if it is Mr. Carson, Mr. Wulf who would want to comment on that.

Mr. CARSON. Yes. This is an issue that goes certainly beyond this particular bill. And let me make clear that we have no particular

position whether such a provision is appropriate in database legislation. I think we would need to hear the case to be made for that. But to the extent that you are referring to a controversy that is very much in the air with respect to section 512 of title 17, which does have the process that you described, we are quite familiar with that and we think most of the complaints that you have just described frankly have very little to them when you peek beneath the hood.

The fact of the matter is that the types of problems that people have claimed exist in that respect first of all are to some extent obviated by the protections that are built into section 512 of title 17, some of which are not in this discussion draft, and we think that if you continue to consider such a provision in this discussion draft you need to consider including such provisions—in particular provisions that require that before you get the subpoena you file something with the court that lays out the basis for your need, including identifying what work is being infringed and where you can find the infringing material.

Beyond that, though, what you have described, if you think about it, isn't so very different from what happens every day in civil litigation. Subpoenas are issued by attorneys in civil litigation without any judge reviewing them. They can seek information about third parties. The third party has no right even to be aware that it is being sought from the person to whom the subpoena is directed until after the fact. So what you are describing is simply a matter of fact, every day in civil litigation in this country.

Mr. KUPFERSCHMID. If I could just supplement what David said, and I agree with everything that he said. To give some context to this, the subpoena provision that is in this bill was drafted at the very outset when the negotiations over this draft legislation had first started. As we all know, a lot has changed over that period of time. But at the time when this provision was drafted the ISPs were supportive of this exact provision that is in the bill. Like I said, I know a lot has changed in the legal regime certainly within those 3 years, but this remains to be an important provision to database producers. We will be, willing to discuss changes to it or to address their concerns, as I mentioned before, if that is necessary.

Ms. SCHAKOWSKY. Did you want to say anything?

Mr. WULF. I am not a lawyer and so I would be happy to respond in writing if that would be all right. I must admit that what I just heard, however, as a layman scares me.

Ms. SCHAKOWSKY. It is disturbing, isn't it?

Mr. WULF. Yes.

Ms. SCHAKOWSKY. Thank you, Mr. Chairman.

Mr. STEARNS. Gentleman from New Jersey, Mr. Ferguson.

Mr. FERGUSON. Thank you, Mr. Chairman. I am not a lawyer either, thankfully. We are beating up on lawyers a good bit today and I usually will jump right in there. I have the AT&Ts of the world in my district. I also have Reed Elsevier, Martindale Hubbell in my district. They obviously are coming at this issue from different sides. I think some very good points have been made quite articulately by our panel today, and I appreciate all of your testimony. My concern is with this chilling effect that we have heard

talked about and actually if I am not mistaken it has been cited on both sides of this debate and this conversation. And to Mr. Donohue I wanted to kind of engage you a little bit on this.

As I have said, I wholeheartedly agree with you and the Chamber on the need for tort reform, for less litigation, for all of the above and you know I have been a loyal foot soldier in that battle and will continue to be.

Mr. DONOHUE. Thank you, sir.

Mr. FERGUSON. And I certainly can appreciate the desire to stay away from increased litigation, increased frivolous lawsuits, et cetera, in the future. And I agree that frivolous lawsuits, litigation trial lawyers, et cetera, have had a chilling effect on business job creation, et cetera, in a whole host of areas. My concern is that—and the way I am seeing this kind of break down a little bit and also with what I have heard in our own office, from folks in my district on various sides of this, is that the folks who are somewhat in the database business, like the AT&Ts of the world, but who have many other business interests are not as interested in this legislation or certainly not supportive of it because they see it as more of a litigation magnet rather than a protection for their business. And the folks at, say, Martindale Hubbell, for instance, and Reed Elsevier, this is their entire business. So they are much more interested in these protections that they would be afforded and their products would be afforded under this legislation. And my question, I guess, is what about the chilling effect on the database collection and management companies? What about the—I guess we are talking about—we have discussed a little bit about a problem that may or may not exist. But some of what we do in the Congress, as you know, is trying to preempt problems that may exist in the future. Maybe we see examples of now, but may become much worse. And there are companies who would probably be doing a lot more investment, a lot more work, a lot more with regard to database collection and providing products to their customers, but maybe aren't doing so today because they know they don't have the protections.

What are your thoughts on kind of the opposite kind of chilling effect?

Mr. DONOHUE. Well, first one sentence to say thank you for your appreciation of the legal questions here and the possibility that this draft would create far more legal difference of opinion and therefore far more cost and, by the way, in those circumstances much less likely for people to invest in database expansion. Second, I think you have to look at the record, and the record is that the database business and the products they are producing become more sophisticated, more expansive and more helpful every day. And the question is, and you could ask that sixth grade class about it because they even use it, the question is what benefit would we get in size and scope to encourage investment, to expand databases or to make us more comfortable with this legislation? And I think there have been some good arguments made here. But I think that the potential, talking about potential difficulty is far more expansive than the potential benefit.

If somebody came back and brought a lot of these extraordinary people that run this business in here and they laid out a whole se-

ries of economic and intellectual property and business difficulties, I would listen because you know what, those people would be my members. But that is not what is going on right now. And I understand that you have a split in your district. But it is pretty clear that we ought not fix it if it is not broken.

Mr. FERGUSON. But what we are hearing from a number of folks, 70 some folks or however many, even it was five, that say that there is a problem.

Mr. DONOHUE. Right. Then I would want to dissect the problem a little more carefully and I would recommend that before you passed a piece of legislation that will affect the information base of an information based economy that is absolutely dependent on it for economic expansion and the creation of future jobs in this country.

Mr. FERGUSON. Very quickly, Mr. Chairman. I know my time is up. You talked about the vague terms of the draft bill and the excessive penalties. I think there probably is going to continue to be a very healthy debate on that. I think probably a lot of us who maybe agree with some points on one side would agree with other points on the other side. If some of those issues were resolved, if this were narrowed and sharpened a little bit, could you see an instance where you could be supportive of this bill?

Mr. DONOHUE. You know I testify a lot here and I like to be very clear, not, you know, take some of the points that others do. But allow me the protection of saying I would have to see it and talk to my members. But from everything I see now, I am opposed to and it would have to be significantly changed.

Mr. FERGUSON. I know my time is up.

Mr. DONOHUE. Thank you very much.

Mr. STEARNS. I thank the gentleman. Mr. Boucher, the gentleman from Virginia.

Mr. BOUCHER. Well, thank you very much, Mr. Chairman. I want to thank these witnesses for being with us today and sharing your views on the subject. You have presented to us well prepared and thoughtful testimony.

We find ourselves, I think, at a fairly curious juncture here. The bill that is before us in draft form is said to be a compromise between the Commerce and Judiciary Committees. I have the privilege of serving on both of those Committees. I may be the only Member here who has that opportunity, and I have had some conversations with many Members of the Commerce Committee in particular about this measure, and I don't believe there is a single Member of the Commerce Committee who is prepared to support this bill. And so I find it very curious that it is being presented today as a compromise measure with the two Committees participating and putting forward this consensus draft. I really find it to be very much to the contrary.

I oppose this measure also, and I want to commend Mr. Donohue in particular for his statement today. I think it was compelling. I think it was incisive. I think he hit all the right points. The remedies that are available under copyright, under trespass, under misappropriation, under contract law itself have proven successful in the cases that have been litigated in protecting databases where there were genuine harms that would arise from the taking of facts

within that database. To legislate at this point, I really think, in the absence of a clearly delineated problem that requires a solution, is simply mischievous and would create an anti-competitive effect, potentially locking away facts that are available to the public today and, at a minimum, requiring that people pay for facts that are free and available for public use presently.

And then, as many Members and some of the witnesses have indicated, there are very substantial constitutional concerns about this measure. So put me down among the opponents. I think the best course we could take is to put this aside and go on to our more urgent legislative business.

Having said that, I just want to ask a couple of questions and, Mr. Carson, I am going to direct these to you. First of all, I am a little bit perplexed by your answer about the subpoenas. What the bill authorizes is subpoenas being issued before any lawsuit has been filed. Now, you talked about the normal discovery process in which lawyers issue subpoenas in civil litigation. But you know there is a fundamental difference between that kind of circumstance and the circumstance presented by this bill. In the circumstance you cite a lawsuit is pending. And in that pending litigation, if the person to whom a subpoena is directed believes that there is some impropriety in that discovery, he has the opportunity to go to the judge before whom that case is pending and seek an order that protects him from this abusive and improper discovery. That happens all the time in our litigation. Under the provisions of this bill that could not happen.

Mr. CARSON. Not at all true, Mr. Boucher. Not at all true. And again I am focusing primarily on what we find in section 512.

Mr. BOUCHER. What is not at all true?

Mr. CARSON. It is not at all true that someone under this bill would be at a disadvantage with respect to someone who is the subject of a subpoena in pending litigation.

Mr. BOUCHER. How does a person under this bill go to a judge and say, Judge, protect me from the issuance of this subpoena or from having to comply with the requirements of this subpoena? How do you do that? You would agree, would you not, that you can do that this regular civil litigation?

Mr. CARSON. You can do it under 17 USC section 512 and although I haven't studied the provision in this discussion draft and I am certainly not here to suggest that it needs to be in here—we have no view on it—assuming that this does track what is in section 512, the subpoenas issued pursuant to section 512 are subject to the Federal Rules of Civil Procedure. In fact, under section 512 people have gone to the District Court here in Washington to seek protection.

Mr. BOUCHER. But you would have to go ab initio. You would have to initiate your own proceeding before the court while there is no judge sitting overseeing that case.

Mr. CARSON. That is not how it works, Mr. Boucher.

Mr. BOUCHER. Well, you and I have a difference of opinion about that and it is obviously something that would need to be examined. I would continue this discussion with you except that I have one other question of you and my time is almost up.

I find another series of provisions in this bill to be curious and I wonder how they can coexist. Perhaps you can help me with this. One of those is the requirement of time sensitivity before the protections this bill would extend to database creators would apply. The other is the fact that this bill is made retroactive. Can you give me an example of a database that contains facts where the protection of those facts is required because of their time sensitivity, where that database is already in existence?

Mr. CARSON. Mr. Boucher, I actually hadn't studied the bill to the point where you saw the retroactivity provision. We have already expressed our concern about the vagueness of the time sensitivity provision.

Mr. BOUCHER. I appreciate that you are not prepared to answer the question.

Mr. CARSON. No, Mr. Boucher. That is not exactly what I said, sir.

Mr. BOUCHER. Well, go ahead if you have an answer.

Mr. CARSON. I have said already we have problems with the time sensitivity provision because we think it is ambiguous and we have already expressed that if the time sensitivity thing carries over for a long period of time there may be problems.

Mr. BOUCHER. Thank you, Mr. Carson. Thank you, Mr. Chairman.

Mr. STEARNS. Ms. Baldwin. Yes.

Ms. BALDWIN. Thank you, Mr. Chairman. Thank you all for being here today. I appreciate your testimony. As the people have been asking questions before, there has been disclosure, so I am a lawyer, but much more importantly, I am the granddaughter and niece of scientists and both belong to one the organizations that Mr. Wulf is representing today. I also represent a major research university, the University of Wisconsin, Madison campus, and so have very strong concerns about the impact of this on the conduct of science.

But before—just as a—given the draft in front of us, there is a question of course of the need for this legislation. If there is need for additional protection for digital databases, I would think that those could ultimately benefit universities and university research efforts both by providing greater incentives to create databases and providing additional protection for university created databases. But obviously I have heard a lot of very generalized concerns from the scientific community about the potential harmful effects of overprotection of databases.

Mr. Wulf, if you can, somewhere between the sixth grade level and the post-doctoral level, elaborate on some of your concerns about the overprotection of databases and its impact on the conduct of science, and especially I want to tease out some details from you, if you can identify some real or hypothetical examples of what might be in store if this legislation were to pass as is.

Mr. WULF. I think something—let me address an issue that you mentioned and that is the value potentially to universities and to researchers of having protection in databases. I have spent about two-thirds of my career in academia and something that is kind hard for people to understand sometimes is that the real motivation, the real inducement for academics to do research is not financial. It is rather peer recognition, and that tends to argue in favor

of as much free flow of information to your peers as possible. So I am not at all sure that in fact additional restrictive legislation here would be beneficial.

The term "chilling" has been used a couple of times here, and I think that is another point that needs to be made. Academics tend to be risk averse. And we will of course never know what they decide not to do because there might be a potential litigation. But academics will, by and large, stay as far away as possible from the potential of litigation. So again we are talking about lost opportunities. We are talking about opportunity costs. Those are very hard to measure. But one should never make the mistake of thinking that because they are hard to measure they are not real. They are very real.

I think some of the kinds of issues that we worry about include access to Government generated information. We worry about access to information generated by sole sources. Much of the scientific data that is collected is observational and it happens at a moment in time with an instrument being available. It is not reproducible. And if that kind of information became inaccessible because of excessive protection, it could have a tremendously chilling effect on the conduct of research. We typically—I shouldn't say typically. Often, often enough to be scary—cannot predict what the use of a database will be. The existence of the ozone hole over Antarctica was verified using databases which had been collected for an entirely different purpose at an entirely different time with no anticipation of this kind of use.

So if access to information that you could not predict would be needed, were restricted, we might not still know about the existence of this enormously environmentally dangerous ozone hole.

Does that help?

Ms. BALDWIN. Yep. Thank you.

Mr. STEARNS. Mr. Delahunt.

Mr. DELAHUNT. Mr. Wulf, I appreciate the concerns that you articulated. You used the term "risk averse and chilling effect." I will tell you what my concern is, that database providers are for profit corporations, and capital oftentimes is risk averse. And I thought Mr. Donohue's observations were correct. You know, his judgment there is not a problem. We hear that there was a case, I think it was the Schoolhouse case and other cases that Mr. Kupferschmid referred to or alluded to during the course of his testimony. It is my belief that there is a problem. But clearly there is a perception among the database community that there is a problem. I know if I were to invest or I know if I were, you know, corporate management and I had capital to invest, I would be looking for new opportunities to diversify. And the concern that I have is all of these great achievements that really have come about because of database and access to information we will not maintain that here in the United States, we will not maintain the level of current data to give us a competitive advantage. And the kind of research that the academies have been interested in will be provided by our, you know, European data providers, data companies. That is the concern that I have.

Now, I think that is a potential problem that impacts exactly what you are talking about because I sense that—and we are only

starting to see in the aftermath much of what was accomplished in terms of the advances that I and others have alluded to was based on work that was protected prior to the Fist case. Now we find ourselves in uncharted waters here, and what I am concerned about is the potential for, again, these data—the database industry diversifying and not investing the kind of resources that are necessary to maintain that absolute current state of data and access to information that provides our scientific community and our economy the kind of advantages that we have enjoyed. I don't know. Would you care to comment or anyone for that matter?

Mr. WULF. Well, I could only repeat myself. We clearly have an innovation system in this country that works. It is an innovation system which is fed by the basic research done in our academic research universities. We just need to be very careful that we don't destroy that in the process. The point I tried to make in my oral testimony was——

Mr. DELAHUNT. Right, and just let me interrupt because we don't have a lot of time. But I agree with that. And I think you testified and it was my understanding that this particular draft that is before us was generated as a result of the various groups whom you testify in behalf of and others sitting down and trying to develop a consensus so that the concerns you express are addressed. Now, presumably, fine-tuning is required. But I guess the bottom line question is who is right? Is it Mr. Kupferschmid or is it Mr. Donohue? You know, is there a problem? If there is a perception of a problem, you know what, there is not a lot of folk that are going to be running out looking to invest in database.

Mr. KUPFERSCHMID. If I could interject here, I mean there are actually numbers that bear that out. In 1996, the EU data base directive was passed. Since that time the percentage of U.S.-produced databases has shrunk from 69 percent to 60 percent .

Mr. DELAHUNT. Well, this is the kind of information I think that this Committee needs. And Mr. Donohue?

Mr. DONOHUE. Well, if you take the primary—the company that has really been behind a lot of this discussion, Reed Elsevier, they have made 20 some acquisitions since 1991, spent in excess of $7 billion and a lot more than that. Those are the numbers that I have here, and their operating margins are still annually over 20 percent. So one might understand why they want to protect their product. But they certainly think it is a good business because they keep expanding it every year and they are getting a hell of a return. I might buy some stock now that I am looking at this.

Mr. DELAHUNT. Well, I will take a look myself. But I guess my question is these new acquisitions, are they an effort to diversify to protect themselves from the kind of liability that I presume they fear that they have, so that, you know, if their stuff continues to get pirated or if any of it gets pirated they are out of—you know, they continue to survive. I don't know. Now, I know that you read that Wall Street Journal, you know, religiously. And maybe you can tell us whether those 19 acquisitions were an effort to diversify or are they just simply building on the so-called core business?

Mr. DONOHUE. Well, Congressman, if we were on their board we would have encouraged the acquisitions because they did diversify their fact base to attract a broader group of fact users. But that

is not the question here and, as my colleague said, this isn't an argument about fact. This is an argument about structure and access to that fact. And what I am having difficulty with is the question of where is the problem? Now——

Mr. DELAHUNT. Well, there was a problem in that Schoolhouse case. 74 percent, that web of laws that you were referring to in your testimony certainly didn't protect that individual.

Mr. DONOHUE. Well, then maybe we ought to go back and look at the case and maybe find out, maybe it didn't deserve to be pro-investigated. Look, I think this joint Committee has done a very useful thing here, because you are having what amounts to a colloquy on the subject, which—and I want to congratulate you on that and tell you I think that is essential and I think you have learned a couple of things, that we are worried about the litigation side of this, that we are very concerned about finding the problem before we get the solution, and that there is legitimate concern on the part of some people about going forward that they protect their assets. Well, what role should the Committee and the Congress have in that and what is in place to do it now, and I think it has been a good discussion and I appreciate the opportunity to participate.

Mr. KUPFERSCHMID. If I could respond. I think Mr. Donohue has done a marvelous job sort of redirecting the questions and staying on script. But this isn't about you know, one company. This isn't just about, Reed Elsevier. There are lots of other companies, a lot of mom and pop companies like Schoolhouse, like Berkshire Publishing, like Carfax, other smaller database producers that are having problems with database piracy. These are companies that really put all their money, all their investment into this one database and if this database is pirated they are out of luck and they are out of business. And so this is not just a one company issue. Like I said earlier, there are lots of members of CADP and they are all supportive.

What I have heard today is that I honestly don't think we can come up with enough examples to ever satisfy Mr. Donohue. He refers to the occasional case. We have got lots and lots of cases and there are a lot of other cases that never make it into the courtroom because the database producers are worried about precisely what you are talking about, that people aren't going to invest in the company or invest in the database if they know how vulnerable the database is. And, you know, there are thousands of realtors out there worried about their databases being on pornography sites. It is a totally different type of concern here. But there is definitely a problem here that needs to be addressed.

Now, if Mr. Donohue doesn't want to recognize that, no matter how many examples we give, I just feel that we will never be able to satisfy him.

Mr. STEARNS. The gentleman's time has expired. We certainly are at a consensus we are not going to have a second round, and we are going to let Chairman Smith close. Oh, okay. Yes, okay. Someone has just come in. Ms. Lofgren, we welcome your questions.

Ms. LOFGREN. Thank you and apologies for my lateness. It is United Airlines' fault. Let me just ask, I guess Mr. Carson or who-

ever wants to answer it. This is not the first time that we have visited this issue. And when last we passed something through Judiciary, I filed a lone dissenting view that in my judgment the measure we passed was—didn't meet the constitutional standard. And actually I was sort of a fan of the sweat of the brow doctrine. I thought it was a very nifty little doctrine that served us well for many years. But we no longer have that available. And I just—I am still not getting how we can create a property right out of something that cannot be copyrighted. And how we are—how this solves this problem. Can anyone answer that for me?

Mr. CARSON. The answer is that no one is proposing that you do that, Congresswoman Lofgren. This is not a property right. This is based on an unfair competition, misappropriation scheme which deals with wrongful use of someone else's database in competition with them. It is not an exclusive rights model such as you find in copyright.

Ms. LOFGREN. Well, I understand that. But ultimately I think it dodges the question, which is in order to protect something you have to have a property right that cannot be created, that I can see. I mean how do you get past that?

Mr. CARSON. You get past it by not giving anyone a property right. This bill doesn't create any property rights whatsoever. It is a right against unfair competition by people who use your material in the same realm that you are using it in direct competition with you in a way that threatens to destroy your ability to continue in business.

Ms. LOFGREN. Well, I think that is a great creative answer, but I don't think it really answers the question. I don't know if anyone else wants to address it but I think that is the meat of what we face here.

Mr. KUPFERSCHMID. Well, I will jump in here. I mean the bill itself in the prohibition itself, forgetting about the exceptions or exclusions in the bill, the bill itself has 10 requirements that any database producer must meet before their database even gets protected here. I mean that pales in comparison to previous legislation that has been out there, where legislation would prevent use or extraction. It covered the potential market here. It has got to be the functional equivalent in the same market. I mean that is a long ways from potential market or related market that were in previous bills. It doesn't cover use of data or information or accessing a database. All it does is cover making available a database in a way that causes commercial harm, and then commercial harm is even defined by a very high standard such that it has got to substantially threaten the incentive to produce the database to begin with. Along with lots of

other—you know, there are nine other requirements here that I haven't even mentioned, so there is a very high standard in the bill. Then when you do include the exceptions and the exclusions here, it would be pretty difficult to prove that there has actually been a violation here.

Mr. DONOHUE. So what we have here then is a limited experience of a problem and a piece of legislation seeking to solve that limited problem which has 10 standards and therefore is probably not going to help very many people, and we have clearly a divided

industry, database industry, looking at the legislation. I still think we have a solution looking for a problem, and I hope we can be very careful in what we do.

Mr. KUPFERSCHMID. With, these 10 criteria here we intended to try to attempt to address the Chamber and other people's concerns. In fact, when this whole process started back in 2001 the Chamber and the Libraries and the University Committee and others provided a document to both Committees, and I will quote exactly from that document in which they said they would support a true misappropriation bill. For example, one which closely follows the historic standards laid out in *NBA* v. *Motorola* which would be constitutional and would not stifle innovation, would not impede scientific progress and would not ultimately hurt the growth of exciting new database products.

Well, we heard the message 3 years ago, or 2 years ago. Here we are. We have got a misappropriation-based approach. One based on *NBA* v. *Motorola*, and we are still hearing the same message. It is a little frustrating because I am not sure what else we can do.

Ms. LOFGREN. I am not sure there is anything you can do. Mr. Wulf, you looked like you wanted to say something.

Mr. WULF. Well, I just—since you weren't here earlier, let me repeat a little bit of what was in my oral testimony. It is estimated that half of the growth in the GDP is due to the innovation system, the advances in science and technology, two-thirds of the growth in productivity due to the advances in science and technology. I am just a little bit concerned here that a small amount of damage to that innovation system in order to protect some pretty isolated cases where this might be useful or appropriate is not the right trade-off.

Ms. LOFGREN. Thank you very much. And given the lateness of the hour, I will yield back the remainder of my time.

Mr. STEARNS. I thank the gentlelady. Mr. Smith will conclude. I will make just a general comment. Mr. Kupferschmid, I think you have made a very strong argument and passionate. Mr. Wulf is saying, you know, between the Digital Millennium Copyright Act, the Computer and Fraud Act, as well as I guess contract law dealing with trespasses, we should move very slow. So maybe just a comment to you is to more narrowly define, tailor your misappropriation statute and maybe come back again at it is a possibility.

But at that I would look to my distinguished Chairman, thank him again for his hospitality here having this joint hearing and allow him to have the last word.

Mr. SMITH. Thank you again, Mr. Chairman. Mr. Chairman, I may have detected a very narrow thread of agreement here. Admittedly it is probably only a nanometer wide, but it is there. And it is this, that no one denies that there are at least some databases that should be protected that are not protected. Mr. Donohue and Mr. Wulf feel that those are isolated cases and that we don't need legislation to address them because they are not as widespread as Mr. Kupferschmid and Mr. Carson believe.

What I wanted to do, Mr. Carson and Mr. Kupferschmid, is to read part of your testimony or at least refer to part of your testimony and ask Mr. Donohue and Mr. Wulf to reply and then you can respond to their comments. The point here is that the proposed

legislation does erect very high hurdles, very high standards that have to be met before anybody is liable for piracy of databases.

Mr. Carson mentioned in his written testimony that basically the legislation codifies the five elements of the Motorola case. Mr. Kupferschmid mentioned in his testimony that the draft legislation creates a narrowly focused prohibition that applies only if 10 criteria are met. So my question really for Mr. Donohue and Mr. Wulf is that, can you think of any example of someone who would meet all these requirements, all these criteria and still be liable for piracy of databases? In other words, aren't these pretty narrowly drawn and maybe your concerns are unwarranted?

Mr. Donohue first, then Mr. Wulf.

Mr. DONOHUE. Thank you, Mr. Chairman. I have observed during my recent tenure at the Chamber over 6 years the work of plaintiffs lawyers up close, and they can find the history of the world written on the head of a pin, and we are creating a piece of legislation with extraordinary penalties in it that will open up a new retirement program.

Mr. SMITH. If we set aside the penalties and just focus on the criteria though, are there any of those criteria that you feel are too broad?

Mr. DONOHUE. I don't have the 10 criteria here sir but I would just say very specifically, if there is overwhelming evidence of economic and intellectual property loss because of behavior against databases I have not seen it. If it is there we want to see it and we would be helpful.

Mr. SMITH. Okay, good. Mr. Carson, can you respond very quickly or perhaps Mr. Kupferschmid?

Mr. CARSON. Mr. Chairman, I don't have fact patterns at my fingertips, maybe Mr. Kupferschmid does, where I can tell you I know of real world cases that would fall into this. Again there was the INS case, there was the *NBA* case. That is what these are taken from. So certainly there have historically been such cases.

Mr. SMITH. Exactly. Thank you.

Mr. Kupferschmid.

Mr. KUPFERSCHMID. Certainly the goal of this bill was to provide a very narrowly tailored misappropriation free riding type approach to database piracy, something that would address the concerns of the database user community while also providing fairly narrow protection. I think the bill comes pretty close to that mark and the cases that we have described, I think, in most cases although, I don't know all the facts in those cases—would be covered by the draft bill under those instances. I also don't think that any existing uses or the manner in which the database user community, make available databases would be altered all under this bill.

Mr. SMITH. Okay. Thank you, Mr. Kupferschmid. Mr. Donohue, if I could ask a final favor of you. Could you get back to us, take a look at those criteria, because if we move ahead we might well want to narrow the bill some more. I don't know. I don't want to speak for the Chairman, but we would be interested in your views as to which of those criteria you feel are too broad.

Mr. DONOHUE. I would be glad to.

[The information referred to follows:]

CHAMBER OF COMMERCE
OF THE
UNITED STATES OF AMERICA

THOMAS J. DONOHUE
PRESIDENT AND
CHIEF EXECUTIVE OFFICER

1615 H STREET, N.W.
WASHINGTON, D.C. 20062-2000
202/463-5300 • 202/463-5327 FAX

October 15, 2003

The Honorable Lamar S. Smith, Chairman
Subcommittee on Courts, the Internet
and Intellectual Property
House Judiciary Committee

The Honorable Cliff Stearns, Chairman
Subcommittee on Commerce, Trade
and Consumer Protection
House Committee on Energy and Commerce

Dear Mr. Chairmen:

I want to thank you again for the opportunity to appear on behalf of the U.S. Chamber of Commerce before your joint hearing to describe our serious concerns regarding the discussion draft of "the Database and Collections of Information Misappropriation Act." On behalf of the Chamber's three million businesses, I also appreciate the opportunity to elaborate on some of the Chamber's concerns, as well as to address some of the issues raised during the hearing.

The Chamber has been actively engaged on the database issues since 1999. As I testified, we have not seen any "gap" in current law that would imperil the ability of database producers – many of whom are active Chamber members – to protect their investments in databases, nor any decrease in incentives to companies to produce databases. Indeed, the testimony of those more sympathetic to the discussion draft, led by Mr. Kupferschmid, has underscored the Chamber's concerns. Specifically, while Mr. Kupferschmid was able to cite a handful of cases that supposedly demonstrated the "gap" in current law, he did not produce the kind of hard economic data of market failure that Congress should require before enacting a new regulatory regime.

The U.S. Chamber typically objects to congressional attempts to legislate "fixes" to anecdotal "crises." Congressional action must be employed to resolve important national challenges, with methods based on quality data and sound science. For example, the Chamber has been urging Congress for several years to implement a comprehensive national energy strategy - a goal I know is shared by the Committee. Various energy producers and consumers have produced an extensive collection of data and projections demonstrating the need for greater domestic energy production and investment to upgrade energy infrastructure. These industry leaders have presented this information in various venues, including in testimony before Energy and Commerce and several other House and Senate committees in the 107th and 108th Congresses.

Echoing the facts and statistics provided by companies and organizations has been the testimony of key government leaders, including Energy Secretary Spencer Abraham and Federal Reserve Chairman Alan Greenspan, highlighting the energy-related challenges that Congress is working to deal with in the Energy Bill. It has been real-world events from the past several years, such as the recent northeast blackout, the California electricity crisis, gasoline disruptions and price spikes, and skyrocketing natural gas costs, that have helped to provide some insight on the complexity of energy issues. Taken with the hard facts, this anecdotal evidence underscores the need for Congress to act.

In contrast, the small number of database publishers advocating legislation (and the number of signatories to Mr. Kupferschmid's letter, which largely consist of subsidiaries of a handful of companies) could not produce any specific numbers concerning their financial health or the amount of "database piracy" that is occurring. Instead of facts, the proponents offer anecdotes, and cite a handful of cases that indicate why there are problems with current law. Indeed, when confronted with figures demonstrating the robust growth and substantial profit margins that this industry has experienced since the *Feist* decision, the proponents simply argue that there is still a disincentive to produce additional databases, although basic economics would indicate otherwise.

**What is the problem? Where is the "gap" in current law?**
*Proponents' case examples do not stand up to scrutiny*

To make their case that there are indeed areas where broader database protection is necessary, proponents of this draft, and Mr. Kupferschmid in particular, cite a number of cases to demonstrate "gaps" in current law. However, a careful examination of these examples clearly indicates just the opposite – current law *does* protect investments in database production and distribution.

*Schoolhouse v. Anderson*, 275 F.3d 726 (8th Cir. 2002).

In this case, the plaintiff, Schoolhouse, accused the defendant of copying his database, which was comprised of information regarding various school districts, such as academic degrees of staff, class size, sports offered, and other publicly available information. Defendant Anderson directly competed with the plaintiff, eventually forcing the plaintiff out of business. However, the district court found that there was no evidence of copying, and noted that Anderson gathered the information in his database independently.

Mr. Kupferschmid asserted during the hearing and in his written testimony that this is an example where a company had a substantial portion of its database copied and then had no legal recourse to protect itself. However, the facts of the case, as found by the district court in granting Summary Judgment for the defendant (so the facts are assumed in favor of the plaintiff), belie that assertion. Specifically, in this case, contrary to Mr. Kupferschmid's assertion that defendant Anderson "admitted [to] copying" 74 percent of the plaintiff's database, the district court noted that the *subject headings* might be similar, but that the actual factual data was separate and distinct. Therefore, this case is anticipated and protected by the draft legislation, which specifically exempts independently-created databases from liability.

Further, Mr. Kupferschmid's contention that this is a prime example of where additional database protection is necessary actually helps to demonstrate the litigation danger that this legislation poses. Specifically, if a database producer can be sued for independently creating a database that the courts found not to be similar to the

plaintiff's database, imagine what would happen if quadruple damages and other vague, undefined terms, discussed further below, were thrown into the mix.

Rather than presenting a case where current law is lacking, this case demonstrates that there is *no* gap in current law. Independent gathering of information is protected under current law, and it would continue to be protected under this legislation.

*Warren Publishing v. Microdos*, 115 F.3d 1509 (11th Cir. 1997).

In this case, Warren Publishing accused Microdos of copyright infringement, and sued to obtain a preliminary injunction to stop Microdos from distributing its product. Although the district court granted a preliminary injunction, that order was vacated by the 11th Circuit, and the Supreme Court denied *certiorari* on the preliminary injunction appeal. Warren Publishing did not pursue the underlying case once the Supreme Court denied the petition.

The facts of this case and the underlying remedies again do not demonstrate the proponents' case that there is a "gap" in current law. Instead, they actually underscore the strong protections that current law *already* provides. Specifically, as Mr. Kupferschmid pointed out in the hearing, after Warren Publishing failed to obtain a preliminary injunction, they began using a shrink-wrap license to protect their publications. In other words, Warren Publishing availed itself of protections that *already exist* under current law, in contrast to Mr. Kupferschmid's assertion that further legal protection is necessary. Not a single case has followed the holding in *Warren's* unsuccessful effort to obtain a preliminary injunction.

There are extremely high hurdles that must be overcome to obtain a preliminary injunction, and it is not surprising that the preliminary injunction in this case was denied. However, in this case, Warren Publishing exercised its rights *under current law*; and not only failed to exhaust its judicial remedies but also failed to avail itself of judicial remedies as to the merits of its underlying allegations.

This case does not provide the Committees with an example of a gap in current law. This case reinforces the notion that current law *does* protect these types of databases.

*Ticketmaster v. Tickets.com*, 2003 Cop. L. Dec. P. 28,607 (C.D. Cal. 2003); 2001 WL 51509 (C.D. Cal. 2001); 2000 WL 1887522 (C.D. Cal. 2000).

*Ticketmaster* is the third case that was highlighted as an example of the shortcomings in current law. In this case, Tickets.com sent an electronic "spider" onto the Ticketmaster web site to extract factual information about upcoming shows, (date, time, venue, etc.), and then created a "deep link" to the Ticketmaster web site, providing its customers an opportunity to purchase tickets to shows and concerts that it did not have access to *directly from Ticketmaster*.

Five courts have considered similar cases, where the plaintiff alleges a state common law trespass to chattels claim. The other four found that sending a "spider" onto someone else's network to extract information, even if it is just factual information, and the presence of the "spider" causes no actual harm to the plaintiff's computer, *does* rise to liability for trespass to chattels. *Ticketmaster* was the fifth, and outlier, case.

In interpreting California law, the federal district court specifically noted that it would stick to its interpretation of the common law "pending appellate guidance." Since the *Ticketmaster* decision, the California Supreme Court has indicated that the reasoning in those other four cases was correct, and that extracting information electronically could constitute a trespass to chattels claim. Given the California Supreme Court's decision, the *Ticketmaster* court likely would reach a different conclusion if it were ruling now.

Further, this case is still ongoing. In particular, the court recently ordered that Ticketmaster's contract claim proceed to trial, meaning that at least one of the plaintiff's claims is still viable. It is also worth noting that Ticketmaster's lawyers failed to allege a violation of the Computer Fraud and Abuse Act, enacted prior to this case, and revised in large part to address these types of issues.

Congress should not change fundamental information policy to address a case in which counsel failed to cite significant existing law, in which subsequent rulings of the state Supreme Court suggested a likelihood that plaintiff would prevail, and in which the plaintiff may yet prevail on its still-pending contract claim.

*Skinder-Strauss v MCLE*, 914 F. Supp. 665(D. Mass. 1995).

Mr. Kupferschmid also alluded to *Skinder-Strauss v MCLE*. In this case, Skinder-Strauss accused MCLE, a non-profit corporation sponsored by the Massachusetts and Boston Bar Associations, of copying Skinder-Strauss's directory of attorneys. While the court held that the listing of attorneys was uncopyrightable, it found that the Skinder-Strauss directory as a whole, which contained a variety of different databases, reflected protectable selection and arrangement. The court has left it to further proceedings to decide whether the overlap constituted copyright infringement. Again, nothing about this case suggests a need for new law, but demonstrates how current law can and does protect databases.

These three examples unfortunately represent the best examples that proponents have been able to cite after more than six years of pursuing this legislation, and more than twelve years since *Feist* supposedly stripped away their database protections. None of these examples stands up to further scrutiny, and we suggest there is no need to create a paradigm shift in the nation's historic intellectual property and information protection regime.

This review of the "horror stories" cited by the proponents underscores the Chamber's belief that there is *no* significant "gap" in the law. Given the enormous number of databases being utilized, there is enormous opportunity every day for some kind of infringement. Instead, however, Congress has been presented with a short list of examples where the databases have been protectable under current law.

**The European Union – Another Flawed Experiment in Database "protection"**

In his oral and written testimony, Mr. Kupferschmid indicated a positive experience with the EU Database Directive as justification for the discussion draft.

However, the Chamber's experience with the EU database directive is far less positive. Adoption of the European Database Directive has led to widespread litigation across Europe. In fact, there have already been close to 100 reported judicial decisions under the EU Database Directive since its adoption in 1996, not including pending actions or cases settled before adjudication. These cases often end with inconsistent results, providing little guidance or comfort. Further, the Royal Society, the British equivalent of the National Academy of Sciences, concluded in April 2003 that the Database Directive "is inappropriate for scientific data and we recommend that it be repealed or substantially amended... "

We believe the United States should not undertake steps to move towards an EU style intellectual property system without a far more systematic review of its implications and a far more careful weighing of the serious risks and potential benefits.

**Specific Problems with Discussion Draft**

During the House hearing, a number of Members requested that I provide specific examples of problems with this draft. Given the discussion above, and the fact that even with more than ten years to find examples of shortcomings in the law, proponents of legislation *still* cannot point to any compelling example where current law failed to provide protection to a database, I see no need for such legislation at this time.

However, there are several significant problems that this draft creates that I would like to highlight for the Committees.

**This is *not* a misappropriation statute.**

In the past, the Chamber has been willing to support a "misappropriation" statute along the lines of H.R. 1858 in the 106th Congress, although legislative proponents *still* have not demonstrated any problems created by a shortcoming in the intervening time frame. Proponents of this draft have incorrectly characterized this draft as a simple "misappropriation" statute designed to stop database "piracy." However, the draft bill subtly, but significantly, expands the traditional notion of "misappropriation," and turns the traditional elements of misappropriation on their head.

The five elements of traditional misappropriation law, as articulated by the court in *NBA v Motorola* 105 F.3d 841 (2d Cir. 1997) are: a plaintiff generates or gathers information at a cost; the information is time-sensitive; the defendant is in direct competition with a product or service offered by the plaintiffs; a defendant's use of the information constitutes free-riding on the plaintiff's efforts; and the ability of other parties to free-ride on the efforts of the plaintiff or others would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened. In at least three of these instances, discussed further below, the definitions and elements of traditional misappropriation law have been dramatically expanded.

**The draft bill's "time sensitivity" and "maintenance of database" provisions could create *perpetual protection* for information contained in databases, far longer than current copyright protection**

**Time Sensitivity:** The tort of misappropriation has its origins in the Supreme Court's 1918 decision in *International News Service v Associated Press*, 284 U.S. 215 (1918) ("*INS*"), which provided a very limited-in-time protection for "hot news." Specifically, the Supreme Court held that basic facts could receive *very limited* protection, if free-riding could result in direct competition for a very limited period of time. Since then, courts have awarded relief in what became known as "hot news" misappropriation cases. Subsequent cases have found, for example, that sports scores are "hot" and potentially protected for a matter of minutes.

The discussion draft, however, creates a new definition of "time sensitivity" in the context of this bill, significantly different from the "time sensitivity" that courts have been familiar with for more than eighty years under the *INS* line of cases. Specifically, this draft would potentially require courts to add the concept of "value" to the determination of time sensitivity, eviscerating the traditional understanding of "time sensitive." Specifically, because some information may retain "value" long after the timeliness of the information has passed, this could extend protection *for as long as information holds value*. For example, legal cases that were written years ago still retain "value," and therefore could still receive protection under this regime. Additionally, the draft explicitly protects facts in encyclopedias, even though the lead-time in publishing means that data is generally months old before it reaches bookstores. In short, this prohibition bears only a superficial resemblance to any logical understanding of "time sensitivity," and instead uses a "value" definition that has little or no relationship to current misappropriation law.

**Maintenance:** Traditional misappropriation law has also required that a person seeking protection must invest resources in *the creation* of a database to qualify for its limited protection. However, the draft bill dramatically extends that protection by providing protection if the database producer expends resources to create *or maintain* a database. This extension creates the possibility of perpetual protection for databases, even those that have material that may be decades old, as long as the producer updates any aspect of the database.

**Vague definitions of "database" and "quantitatively substantial part" of a database invite uncertainty and litigation.**

The proposed definition of "database" raises obvious threshold questions of how large a number of items would qualify as a collection of information, and just how much need to be "taken" to invoke the prohibition. That question is made particularly difficult by the proposed inclusion of "compilations" and "collective works," which are traditionally the subject matter of copyright and which will test the limits of how much is a "large number" in the context of such works. The idea that a subset of a "database" can also be a "database" promises further confusion. Moreover, even the most trivial of databases arguably contain large numbers of

discrete items. Even an individual's monthly credit card statement, for example, typically contains hundreds of discrete facts and pieces of information.

Additionally, the determination of what is a "quantitatively substantial" part of an undefined "database" is unclear. Case-by-case, fact-intensive determinations and adjudications of that sort of question are likely to produce little guidance and conflicting results that give no useful direction to businesses. These ambiguities invite costly litigation.

**The draft legislation is *retroactive*, so it would not only provide perpetual protection to new databases, but would cover data that is decades old.**

One key element of misappropriation law is that misappropriation could reduce the incentives to continue to produce databases. However, as we have seen from the phenomenal success of the database industry, such incentives to produce databases are certainly not lacking in the current marketplace. However, the retroactivity of this legislation underscores our concerns that this legislation is not intended to create an incentive to produce databases in the future, but could be used to severely restrict the ability to use facts that are currently in the public domain *on a going forward basis*.

**The "remedies" that the bill proposes will serve to encourage litigation and stifle innovation.**

As discussed above, this draft legislation contains vague terms which will discourage the use of databases for all of the innovative purposes that Mr. Kupferschmid cited in his testimony. It will also encourage litigation, because no party will know how strong a case is until the court has ruled. Further, because those rulings will be very fact specific inquiries, few decisions will provide any certainty for future conduct or cases.

In addition, the monetary awards are substantial, and will encourage frivolous litigation and settlements. The draft not only provides the plaintiff with recovery of

"actual damages," but those damages are based upon *the profits of the defendant*, not the harm to the plaintiff. Additionally, the defendant is charged with the burden of demonstrating that any revenues should not be included as part of the damage award. That creates a strong incentive for a plaintiff to bring a case, even if the chances of success are small.

Additionally, the draft contains what are in effect threats to defendants if they do not settle. For instance, unlike antitrust or other commercial law, this draft awards plaintiffs with up to *quadruple* damages. The determination of whether such an award is justified is based on factors such as ability to pay and "good faith effort[s] to rectify the misappropriation." In other words, if a defendant vigorously defends itself in court and refuses to settle, that may be construed as bad faith, and could result in a substantially higher reward for a plaintiff.

**Conclusion**

On behalf of the Chamber's three million businesses, I want to again thank you for the opportunity to discuss the Chamber's serious concerns regarding the database discussion draft. The Chamber has been an active participant in this process for four years, during which time we have not seen a demonstrated need for this legislation. The examples broached at the hearing – and discussed above – underscore the absence of a "gap" in the law. I reiterate the common-sense observation that, if Chamber members couldn't protect their investment in database production, they would be strongly urging me to fight for new protection. Instead I'm hearing that there is little or no upside for the business community in database legislation, and potentially a significant, anti-competitive downside.

This legislation, if enacted, would combine vague terms with excessive penalties to not only stifle innovation, but to create a litigation nightmare for businesses of all industries. The ten criteria articulated in the draft legislation collectively create a statute that – while superficially resembling a misappropriation bill – bears little actual relation to misappropriation standards business has lived with and courts have interpreted for more than eighty years. To simply focus on the core prohibition of "time sensitivity," for instance, the draft moves from the standard created by the Supreme Court in *INS* and preserved by the Supreme Court in *Feist* – in which facts

The Honorable Lamar Smith
The Honorable Cliff Stearns
October 15, 2003
Page Twelve

in wire stories were "hot" for hours and sports scores are "hot" for minutes - to a standard of retroactivity that ensnares facts in databases that are conceivably decades old.

Clearly this prohibition looks more like the copyright of facts forbidden by *Feist* than the constitutionally sanctioned "hot news" misappropriation with which business is familiar. It is clear that these vague terms, private right of action, quadruple damages, and incredibly expansive subpoena power will encourage frivolous litigation.

In the Information Age, fundamental changes in basic information policy will affect virtually every American, as well as virtually every business, not just those commonly thought of as information companies. The discussion draft creates an environment more fertile for unnecessary litigation than for database production.

Again, on behalf of American business and our three million members, I want to thank you for inviting me to testify and share our serious concerns.

Sincerely,

Tom

Thomas J. Donohue

---

Mr. SMITH. Okay. Thank you again, Mr. Chairman.

Mr. WULF. Let's see. Could I just jump in with a 10-second one here because I would second what Tom Donohue said. I would just like to add to that that we need to look at this in the light of the developments that have happened since the last time we went around this race track. Things have changed. The environment has changed, and so we need to be very careful that we take into account those changes.

Mr. SMITH. Okay. Thank you, Mr. Chairman.

Mr. STEARNS. I thank my distinguished colleague. And with that, we thank the witnesses very much for your enthusiastic, energetic testimony, and we look forward to continuing discussion.

With that, the Committees are adjourned.

[The prepared statement of Mr. Dingell follows in the Appendix]

[The prepared statement of Mr. Shimkus follows in the Appendix]

[Whereupon, at 5:50 p.m., the Subcommittees were adjourned.]

# APPENDIX

## MATERIAL SUBMITTED FOR THE HEARING RECORD

### PREPARED STATEMENT OF THE HONORABLE W. J. BILLY TAUZIN, CHAIRMAN, COMMITTEE ON ENERGY AND COMMERCE

Thank you Mr. Chairman for holding this hearing today. It is always a pleasure to work with my colleagues at the Judiciary Committee. I value the insights the members of your Committee bring to the issues over which we share jurisdiction.

The database issue has been around for some time now. I remember the first bill being introduced in the 104th Congress and reaching a peak of contention in the 106th Congress. At that time there was a stalemate between these two Committees. Each Committee passed its own bill - staking out its own position on the issue. Neither Committee was willing to move toward the other. When I took the gavel at Energy and Commerce and Chairman Sensenbrenner took over at the Judiciary Committee, we decided we would work through this issue in a different way. The two Committees have worked amicably towards a draft bill - and I believe this is a credit to the fine members on both Committees.

So here we are today . . . We devoted 2 1/2 years of resources to get a draft piece of legislation and to get that legislation before the two Committees for a full and fair vetting of the issues. And that is what I expect today - a fair hearing on the issues involved. I do not expect the issue to be less contentious than it has been in the past but I do expect it will take on a new civility due to the cooperative nature in which the two Committees have been working.

We have a distinguished panel of witnesses with significant expertise on these issues. The witnesses on the panel have been active in the database debate over the last several Congresses and are no strangers to those of us who have followed the debate. I look forward to hearing your perspectives on the draft legislation and drawing on your expertise as we talk through the issues before us. I thank you all for being here this afternoon and yield back the balance of my time.

### PREPARED STATEMENT OF REPRESENTATIVE BART STUPAK, SUBCOMMITTEE ON COMMERCE, TRADE AND CONSUMER PROTECTION

I appreciate this joint committee forum to discuss this bill- and want to express my concerns with moving forward with such legislation.

In today's information age, databases are the tools that make vast amounts of facts and information understandable and manageable.

The facts the databases rely upon are public domain. Facts cannot be owned.

But this bill seeks to do just that- to grant the compiler of a database unprecedented ownership rights to facts.

Current law is sufficient to deal with the misappropriation of information or infringement upon creative works.

The broad opposition to this bill, ranging from consumer groups to database producers themselves, is very telling.

If a need existed for such legislation, surely multiple database producers would be clamoring for such a bill, rather than expressing major concerns.

To move this bill forward would be to move the flow of facts and information backwards, and would disadvantage consumers.

I certainly believe that this hearing is informative, but do not believe that any further steps should be taken on this bill to advance it.

Thank you.

PREPARED STATEMENT OF REPRESENTATIVE BARBARA CUBIN

Thank you to both Chairmen for their commitment to working together on this issue and holding this hearing today. It is important that the discussion continue, as we seek to determine what, if any, remedy would be most appropriate.

I would also like to thank the distinguished panelists that have joined us today. Your testimony is valuable and essential in furthering the debate on database management and protections.

A well balanced policy in this and every realm is an important goal for Congress. While this debate has historically dead ended in its search for that balance, it does not mean we should forego our efforts altogether.

One's creation, original or compilation, should certainly be afforded rights and protections. It is equally important, particularly in today's world of ever changing and exponentially growing technological advances, that these protections not inhibit further development and available options.

Again, I thank the panelists and am certain that today's testimony will further illuminate the path that this legislation must take in a timely manner.

I thank the Chairman again and yield back the remainder of my time.

---

PREPARED STATEMENT OF REPRESENTATIVE GENE GREEN

Thank you to our Chairmen and Ranking Members for holding this hearing on the need for expanded legal protection for databases. I appreciate our witnesses coming before us today to give us their views on this draft legislation.

This country has a long-honored tradition of considering factual information part of the public domain. In fact, the ability of scientists and researchers to have unrestricted access to this public information has contributed to the tremendous innovation on which this country's economic strength rests.

For several years now, the database industry has come to Congress with their concerns about database piracy and the effect that it will have on the industry's willingness to invest sufficiently in new products.

While I understand their concerns, my initial thought is that these concerns seem a little premature considering that the digital age has only contributed to the proliferation of databases. And, to date, I have not seen any real evidence of investment in databases being stymied.

Even if we assume that this threat to the database industry is real, I have questions about the necessity of enacting such a broad piece of legislation to protect them. In the past, the Energy and Commerce Committee has approved legislation to narrowly address this very issue, and I question why we are not taking a similar approach today.

Again, I thank our witnesses for appearing before us today. I look forward to your testimony and the light that it will shed on this important issue.

---

PREPARED STATEMENT OF THE HONORABLE TED STRICKLAND, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF OHIO

Thank you Mr. Chairman. Briefly, I would like to express my reservations regarding the Database and Collections of Information Misappropriation Act. I have heard from a number of interested parties about their concerns with this legislation and I hope this hearing encourages Congress to continue thinking about legislation that would create broad new protections for databases.

In the 106th Congress, I supported H.R. 1858, the "Consumer and Investor Access to Information Act of 1999." This bill was more narrowly written to create new protections against the selling or distributing of duplicated databases in interstate and foreign commerce. While H.R. 1858 would have offered new legal protections for database owners, these protections would not have limited the American public's access to information. It may be that additional protections for database owners are worthy of pursuit. However, facts that are part of the public domain should remain so and I hope we are careful to ensure we preserve the public's access to data and information and avoid unintended consequences as this debate continues.

I think the Database and Collections of Information Misappropriation Act of 2003 is too broad. I fear it would change our current information policy to a point where we could stifle innovation, hamper scientific progress and get in the way of development in the electronic commerce marketplace.

Again, thank you Mr. Chairman.

---

# CARDOZO

BENJAMIN N. CARDOZO SCHOOL OF LAW • YESHIVA UNIVERSITY

Justin Hughes
*Assistant Professor of Law*

(212) 790-0260
FAX (212) 790-0205
E-MAIL hughes@ymail.yu.edu

September 23, 2003

The Honorable F. James Sensenbrenner, Jr.
Chair, House Committee on the Judiciary
2449 Rayburn House Office Building
Washington, D.C. 20515

The Honorable W.J. "Billy" Tauzin
Chair, House Committee on Energy and Commerce
2183 Rayburn House Office Building
Washington, D.C. 20515

Dear Chairman Sensenbrenner and Chairman Tauzin:

On September 2, 2003, your respective committees released a new proposal for extra-copyright protection of databases ("the Proposal"). This Proposal is the latest effort in our country to respond to the Supreme Court's elimination of "sweat of the brow" copyright protection of large, comprehensive databases in *Feist Publications v. Rural Telephone*, 499 U.S. 340, 111 S.Ct. 1282 (1991) and the 1996 creation of a *sui generis* intellectual property right in the European Union to protect investment in such comprehensive databases (The "EU Database Directive"[1]).

The present Proposal is clearly the result of enormous hard work. Those of us who have worked on this issue in recent years know the difficulties committee staff have faced in trying to reconcile sharply divergent perspectives on what kind of extra-copyright protection, if any, is needed for databases.

Today the relevant subcommittees (the Subcommittee on Courts, the Internet, and Intellectual Property and the Subcommittee on Commerce, Trade, and Consumer Protection) will be holding a hearing on the Proposal. It also seemed like an appropriate time for neutral parties to offer some initial impressions on the Proposal. I believe that the Proposal creates a cause of action that is clearly more

limited to genuine misappropriation than some of the prior proposals that came before Congress. At the same time, the Proposal has not completely solved many of the problems in protecting database investments against misappropriation without disturbing the free flow of information necessary for science, education, and a vibrant 'information economy.'

The analysis that follows is a preliminary review of the Proposal, divided into a discussion of the Proposal's "basic prohibition" (Part I); a discussion of other important aspects of the Proposal (Part II); and consideration of whether this Proposal is "comparable" to the EU Database Directive, such that under Article 11(3) the European Union will grant its *sui generis* protection to American database makers who would not otherwise qualify.

Parts I and II conclude that the Proposal is a better approach than prior bills before your respective committees, but still has some elements that are likely to trigger mischief and abuse. Part III concludes, based on a recent presentation by an EU official, that the Proposal would not be accepted by the European Commission as "comparable" to the EU Database Directive.

The analysis that follows only occasionally touches upon the constitutionality of the Proposal, although, as you know, that is a central consideration in drafting legislation in this area. A number of scholars and experts have written about the constitutionality of database protection.[2] To the degree the Proposal embraces a genuine misappropriation approach, it increases its chances of passing constitutional muster. The Proposal's constitutional "profile" is also an improvement over some prior proposals in a number of areas, such as elimination of criminal liability.

To the degree the Proposal moves to embrace a genuine misappropriation approach it also improves the *theoretical* case justifying additional federal law. We must be cognizant that any database protection law is a proposal for *additional government regulation of the information economy*. One of the basic problems with the very strong extra-copyright database protection in the EU Database Directive is that there is no evidence that it has improved Europe's information industries; the only thing we know for certain is that the Database Directive has increased litigation in Europe, principally over databases that already existed before the "incentive" of the Directive was created. Similarly, in the U.S. there is no evidence of market failure – and, therefore, no need for anything resembling the additional regulation of the Directive. In the absence of such empirical evidence, the present Proposal offers a more limited and narrow protection that is more easily justified by the basic economic case for investment in information products.

All of us who have worked on database protection appreciate the time and consideration you and your respective staffs' have put into this project – as well as your willingness, signaled by this Proposal, to continue to listen to interested parties and the public on this issue.

Cordially,

Professor Justin Hughes

Attachment – Draft Analysis

**ANALYSIS -- Draft**
Justin Hughes, Cardozo School of Law

Introduction
I. Analysis of the Proposal's Basic Prohibition
II. Analysis of other elements of the Proposal
III. Is this Proposal "comparable" to the EU Database Directive?
Conclusion

**I. Analysis of the Proposal's Basic Prohibition**

The basic prohibition of the Proposal is contained in section 3. Section 3 is considerably more complex than the basic prohibitions that were in HR 354 and HR 1858. It appears that the drafters of the Proposal wanted to limit the "footprint" of the basic prohibition instead of creating a broader prohibition with lots of exceptions, exemptions, and limitations. Generally speaking, that approach should produce a simpler law, making it easier for citizens to understand what is prohibited and what is permitted.

But a "basic prohibition" can itself become very complex and there are *lots* of factors built into section 3. I would recommend against any additional complexity in drafting this basic prohibition. The basic prohibition is broken into its operative clauses for easier discussion below:

"Any person who
A. makes available in commerce to others
B. a quantitatively substantial part of the information
C in a database generated, gathered, or maintained by another person,
D knowing that such making available in commerce is without the authorization of that person (including a successor in interest) or that person's licensee, when acting within the scope of its license,
*shall be liable for the remedies set forth in section 7 if –*
E. (1) the database was generated, gathered, or maintained through substantial expenditure of financial resources or time;
F. (2) the unauthorized making available in commerce occurs in a time sensitive manner and inflicts injury on the database or a product or service offering access to multiple databases; and

G. (3) the ability of other parties to free ride on the efforts of the plaintiff would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened.

The meaning of many of these clauses depends on definitions explicitly laid out in the bill. Your staff should be commended for their genuine effort to clarify these elements through the section 2 and section 3(c) definitions. Let me address each of the elements of the section 3 basic prohibition individually.

**A. makes available in commerce to others**

This is a substantial improvement over the broad proposal in the original HR 354. As originally proposed, H.R. 354's basic prohibition consisted of three basic elements, imposing liability on any person who "extracts or uses in commerce" all or a substantial part of a database so as to cause "harm" to the "actual or potential market" of the database creator. Instead, the Proposal's prohibition of "mak[ing] available in commerce to others" appears to be similar to the scope of HR 1858, which would have created liability for those who "sell or distribute to the public" a database belonging to another.

But, unlike HR 1858, the new bill adds a definition of "making available in commerce to others" [section 2(14)] which tries to clarify which distributions trigger liability and which distributions would not. It appears that the section 2(14)(B)(i) definition is intended to ensure that distribution within a family or small circle of friends or social acquaintances does NOT trigger liability ("a family and its social acquaintances"). That is a laudable goal, but does the "social" modifier in front of "acquaintances" mean that *work* colleagues are excluded – with the intent that they fall under §2(14)(B)(ii)? If so, that points needs to be clarified and discussed.[3]

Section 2(14)(B)(ii) creates liability when the distribution "extends beyond . . . those who could reasonably anticipate to have a database made available . . . to them . . . without a customary commercial relationship." This is also worth discussion. If the intent is to protect intramural distributions of data within non-profit and educational institutions as well as many extramural distributions of data among researchers, educators, and scientists, then the legislative history could make this clear – and make clear that Congress intended to protect what were *customary*, non-commercial distributions of data *prior to the passage of any bill.*

But section 2(14)(B)(ii) may not help educators, researchers, and scientists enough because of the present construction of section 2(14)(A) and (B). Subsection (A) says "a substantial number of members of the public" and subsection (B) says "a number of persons that extends beyond" either "a family and its social acquaintances" [(B)(i)] or the group that could reasonably anticipate getting the database "without a customary commercial relationship" [(B)(ii)]. The question is: what is the interaction of provisions (A) and (B)?

There is a danger that under (B) "a number" could be *one person*. In other words, does distribution to one (1) person "beyond a family and its social acquaintances" count as "making available in commerce to others"? Does distribution to one (1) person "beyond . . . those who could reasonably anticipate" getting the database "without a customary commercial relationship" count as "making available in commerce to others"? If so, then (A) is of limited importance and the Proposal could have a significant chilling effect on simple data exchanges among citizens where commercial databases are involved because no one will know when giving a single copy to someone will move them "beyond . . . those who could reasonably anticipate" getting the database "without a customary commercial relationship."

To cure this serious problem, I suggest the Subcommittee consider collapsing § 2(14)(A) and (B) to read, simply, "a substantial number of persons beyond . . ." followed by the (B)(i) and (B)(ii) group descriptions.

A final point that warrants some emphasis. Because the Proposal can only be passed under Congress' general Commerce Clause power, "making available in commerce to others" is constitutionally limited to commercial exchanges affecting interstate commerce – a point correctly recognized by Section 2(2) of the Proposal which provides that "commerce" means "all commerce which may be lawfully regulated by Congress."

Parallel language is used in the Lanham Act. 15 U.S.C. § 1227. *See Larry Harmon Picture Corp. v. Williams Restaurant Corp.*, 929 F.2d 662 (Fed. Cir. 1991), *cert. denied* 502 U.S. 832 (1991) (finding, for purposes of Lanham Act, interstate commerce in a restaurant with interstate customers). Even if the proposal were silent on a definition of "commerce," the limitation is a matter of Constitutional, not statutory law. To the best of my knowledge, there have been no significant tests of what constitutes interstate commerce for purposes of trademark law since the Supreme Court's decision in *United States v. Lopez*, 514 U.S. 549 (1995).

The *Lopez* decision in tandem with existing jurisprudence under 15 U.S.C. § 1227 suggests that there will be some – possibly many -- database distributions which would not rise to the level of "interstate commerce" that can be regulated by Congress.

**B. a quantitatively substantial part of the information**

The Proposal deserves praise for having eliminated the idea that a quantitatively non-substantial, but **qualitatively substantial** redistribution of information would be actionable under a database protection scheme.

The idea that a quantitatively small taking can still be substantial because it takes the heart of a work was enshrined in copyright law in *Harper & Row v. Nation*, 471 U.S. 539 (1985). In that case, the Court concluded that the copying of 300 words from former President Gerald Ford's 450-page biography constituted actionable copying. While the *Harper & Row* reasoning makes sense with literary and artistic works, it makes much less sense with large, comprehensive databases. Yes, there will be *some* databases that have *some* entries that are more rare and harder to obtain than other entries. But the added, miniscule "incentive" would not be worth the risk of vexatious litigation from allowing plaintiffs to claim that quantitatively non-substantial redistribution creates liability. *See* Justin Hughes, *How Extra-Copyright Database Protection Can be Constitutional*, 28 U. Dayton L. Rev. 159, 212 - 213 (2002).

**C in a database generated, gathered, or maintained by another person,**

The important elements of this phrase in the basic prohibition are discussed under "E" below.

**D knowing that such making available in commerce is without the authorization of that person (including a successor in interest) or that person's licensee, when acting within the scope of its license,**

The addition of this knowledge standard is a new and welcome element to the database protection discussions. This phrase appears to move the Proposal away from a strict liability statute. Under the Proposal, knowledge of lack of authorization would be *required* for liability under the statute, in contrast to classical intellectual property laws. *See, e.g. Shapiro, Bernstein v. 4636 S. Vermont*, 367 F.2d 236, 238 (9th Cir. 1966) ("knowledge of lack of authorization is not

required for a violation of the copyright law"); *Bennett v. Flash 1-Hour Foto*, 1994 U.S. Dist. LEXIS 6704 (N.D. Ill. 1994) (same).

This knowledge standard also helps focus liability on commercial enterprises, particularly competitors of a database maker, who will be familiar with the "rules of engagement" and focuses liability *away* from private citizens and individuals who do not understand well-established federal laws, let alone completely new causes of action.

**E. (1) the database was generated, gathered, or maintained through substantial expenditure of financial resources or time;**

The idea of investment in "generating" and "gathering" information seem familiar and relatively non-controversial. The two most interesting points in this clause are (a) what kind of "maintenance" of a database is sufficient to trigger protection under the Proposal separate from "generating" and "gathering"; and (b) what counts as a "substantial expenditure" of financial resources or time that will trigger protection under the proposal.

The proposal clarifies/limits "maintenance" for purposes of section 3. Section 2(13) states that "To 'maintain' a database means to update, validate, or supplement the information contained in the database." This helps eliminate *some* overly broad understandings of "maintaining" a database: simply keeping a server plugged into electricity would no longer count as "maintaining" an online database. But, as I understand it, the Proposal still does not solve the problem of large, comprehensive databases which are maintained [in the section 2(13) sense] over the course of years or decades. Under sections 2(13) and 3, "supplementing" a huge database with a relatively small number of new entries could count as "maintenance" justifying protection of the entire database. This issue would come to the fore if the Proposal were amended to include a definite term for database protection.

**F. (2) the unauthorized making available in commerce occurs in a time sensitive manner; and inflicts injury on the database or a product or service offering access to multiple databases; and**

It's clear that inclusion of the "time sensitive" language here is intended to echo the holding in *National Basketball Association v. Motorola*, 105 F.3d 841 (2d Cir. 1997) which found that New York state common law misappropriation could survive preemption by federal law under certain limited conditions. The

*NBA v. Motorola* analysis was inspired by the circumstances and reasoning of the Supreme Court in *International News Service v. Associated Press*, 248 U.S. 215 (1918). In the *INS* case, the Supreme Court recognized a "quasi-property" right to information beyond – but apparently compatible with – copyright law. The Court implicitly was concerned with very "time sensitive" information, noting that "[t]he peculiar value of news is in the spreading of it while it is fresh" 248 U.S. at 235. Court and scholars would later call this a "hot news" analysis, although the two phrases – "hot news" and "time sensitive" were, for different reasons, not the sort of language the 1918 Supreme Court employed.

But the Proposal likewise uses "time sensitive" in a way that would be unrecognizable to courts and commentators. Section 3(c) leaves the "time sensitive" status of data in the hands of courts, instructing only that "[I]n determining whether an unauthorized making available in commerce occurs in a time sensitive manner, the court shall consider the temporal value of the information in the database, within the context of the industry sector involved."

The phrase "temporal value" appears to mean simply "value" such that section 3(c) produces a 'soft' definition that information is "time sensitive" as long as it is valuable. If information is valuable in a particular "industry sector" for 20 years, then under section 3(c) the information is "time sensitive" for 20 years. That is not how any American courts have ever used the phrase. *See, e.g. National Resources Defense Council v. Evans*, 316 F.3d 904 (9th Cir. 2003) (discussing "time sensitive data"); *NBA v. Motorola, supra* (constantly changing sports scores as "time-sensitive information"); *Gustafson v. Jones*, 290 F.3d 895 (7th Cir. 2001) (discussing "time sensitive information" in the context of criminal investigation). In an effort to reflect the *NBA v. Motorola* analysis, the Proposal stretches the notion of "time sensitive" inappropriately.[4]

The result is that, instead of limiting protection, the "time sensitive" concept in the Proposal creates the prospect of permanent protection. If information is "time sensitive" as long as it is valuable, then anytime a plaintiff is economically inspired to sue, they have a cause of action and any limitation created by section 3(c) is illusory. This was one of the Administration's principal criticisms of H.R. 1858 in 1999. As then Department of Commerce General Counsel Andrew Pincus stated:

> "We do not support the basic premise of H.R. 1858 . that a codification of misappropriation principles should provide an open-ended term of protection because common law misappropriation principles do not

impose any fixed duration to such claims. We also believe that legislation must specify the acts that initiate the term of protection."[5]

This concern is as valid today as it was in 1999. An unusual definition of "time sensitive" should not produce the prospect of permanent protection for any database under the statute. There is *absolutely no evidence* that such an economic incentive is needed for the production of any databases.

**G. (3) the ability of other parties to free ride on the efforts of the plaintiff would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened.**

This is another component of section 3 which appears to be inspired by the analysis in *NBA v. Motorola* (105 F.3d at 852) as well as the approach taken in H.R. 1858. The problem with this requirement for liability is twofold: either (a) it will apply in *all or almost all* situations or (b) it will be completely non-transparent: how will a person who did not develop the database know when "the ability of other parties to free ride on the efforts of the plaintiff would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened"? Again, this is a concern that the Administration raised in 1999 in relation to H.R. 1858.

## II. Other elements of the Proposal

There are many other elements of the Proposal that deserve comment. Again, in general, the staff deserve great credit for having worked to craft a legislative proposal that strikes a middle ground aimed at addressing unfair competition without unduly restricting the flow of information in society. At the same time, the Proposal has some elements that should be removed, refined, or adjusted.

### *A. "Discrete" subsections of databases as "databases"*

Perhaps the most important element of the Proposal outside the basic prohibition of section 3 is the Section 2(5)(C) treatment of "discrete sections" of databases:

> "(C)Discrete sections. – The fact that a database is a subset of a database shall not preclude such subset from treatment as a database under this Act."

It is hard to overstate the importance of this a short provision on the scope of the proposed database protection.

This provision in the Proposal is presumably inspired by a parallel provision that existed in H.R. 1858. It was a troubling idea then and it is a troubling idea now. The Administration criticized this type of provision in its 1999 testimony on H.R. 1858:

> ". . . . section 101(1) provides that a discrete section of a database 'may be treated as a database.' We recognize that the intent of this 'discrete section' provision is to protect identifiable subsections of databases from wholesale misappropriations, but we very are concerned that this definition could create liability for insubstantial distributions from databases, particularly in the digital environment.
>
> "For example, the book edition of a national database of hotels might subdivide hotels by state and city; in such a situation, we understand that the intent of section 101(1) would be to create liability when a competitor misappropriated all of the Sacramento, California or Cincinnati, Ohio listings from the national database, even though this might only be a small part of the national database. But in a digitized form, the same national hotel database can have discrete sections organized by state, by city, by neighborhood, by quality rating, by hotel ownership or chain participation, by price, by the availability of particular services (conference rooms of such a size, gym facilities), etc. . so that many, if not most, distributions of material from the database could trigger the 'discrete section' provision.
>
> . . . . .
>
> "Defining a database to include a discrete subset of the database invites database producers to format their products so as to make small amounts of the data appear as 'discrete,' therefore liability-triggering, subsets. We therefore recommend against this approach."[6]

On closer scrutiny, section 2(5)(C) may be more troubling than the provision in H.R. 1858. Section 101(1) of H.R. 1858 provided "However, a discrete section of a database that contains multiple discrete items of information may also be treated as a database." But while "discrete sections" is the title of section 2(5)(C) of the Proposal, it is not in the operative legislative text. Section 2(5)(C) does NOT say "[t]he fact that a database is a **discrete, readily identifiable subset** of a database shall not preclude such subset from treatment as a database under this Act."

The danger of mischief, manipulation, and abuse with this kind of provision is as great now as it was in 1999. For this reason, the "discrete sections" language should be struck for the time being. In many ways, the problem of identifying "the database" for purposes of a misappropriation claim are parallel to the problem courts often confront in identifying the "work" for purposes of copyright law. The 1976 Copyright Act has functioned quite well for almost three decades with **no** definition of a "work." Judges should also be able to make sensible determinations about what counts as a "database" vis-à-vis protection against misappropriation.

If members of the Committees remain concerned that free-riders will appropriate valuable portions of large databases without such a discrete subsections provision, then one recommendation would be to include a provision establishing that the issue should be studied 5-10 years after implementation of the bill by the U.S. Patent and Trademark Office, the Federal Trade Commission, and the Department of Justice, in cooperation with the Copyright Office. Such a study provision could target this as a specific issue, much as provisions of the Digital Millennium Copyright Act do. *See* 17 U.S.C. §1201(a)(1)(C) [study on impact of §1201 on availability of particular classes of works]; 17 U.S.C. §1201(g)(5) [study on impact of §1201 on encryption research]. Five to ten years would be a more suitable time frame then the shorter periods that have been used for follow-up studies in the DMCA. As a general rule, markets do not respond as quickly to changes in the incentive structure created by intellectual property laws as many people in Washington have believed.

**B. *Simplifiying the Proposal: delete some definitions***

The Proposal includes definitions of "Collective works" and "compilations" at Section 2(1) and 2(3), respectively. These definitions are then used in the Section 2(5)(B)(i) definition of a "database" to ensure that what are compilations under copyright law are treated as databases under this law. Section 2(5)(B)(i) states that a "work(s) of authorship" are excluded from database protection, "other than a compilation or a collective work."

Instead of these statutory definition, it would be simpler to provide in Section 2(5)(B)(i) that protected databases exclude any "work of authorship, other than a compilation under 17 U.S.C. § 101." In Title 17. "compilations" are statutory defined to "include collective works," so it is even unnecessary to state

"work of authorship, other than a collective work or compilation under 17 U.S.C. § 101.) (Although this too would eliminate the need for the definitions.)

Including the definitions of "collective work" and "compilation" in the Proposal itself may have been intended to further separate the Proposal from copyright law. But the definitions are taken verbatim from the copyright statute and the relationship between the two forms of protection should be acknowledged, not hidden. Companies that use intellectual property to protect their investments, intellectual property practitioners, and the judiciary are all accustomed to a jurisprudence in which more than one form of intellectual property may protect one creation or work. Software may be protected by both patent and copyright; cartoon characters may be protected by both copyright and trademark; some elements of a technological advance may be protected by patent, while other elements are protected by trade secrecy law. **Some** overlap has come to be expected. *See, e.g.* U.S. PATENT AND TRADEMARK OFFICE & U.S. COPYRIGHT OFFICE, PATENT-COPYRIGHT LAWS OVERLAP STUDY x (1991) (prepared for the House Subcommittee on Intellectual Property and the Administration of Justice, and concluding that "there appears to be minimal overlap between the subject matter of copyrights and utility patents because the statutes make clear that the areas of protected matter in each case are markedly different"); *Mazer v. Stein*, 347 U.S. 201 (1954) (holding that copyright and design patents could protect the same work).[7]

Since the early discussions of extra-copyright protection of databases, it has been understood that any system would almost certainly apply to some works already eligible for (at least some) protection under the Copyright Act. The EU's Database Directive is, for example, expressly drafted with two overlapping forms of protection: the copyright protection of database selection/arrangement and the *sui generis* protection of investment in databases. A statutory reference to Title 17's definitions will only clarify what is already understood by all who have participated in these discussions.

### C. *Clarifiying the Proposal: refining some definitions*

The definition of "Internet Access Service" in section 2(10) states that "[s]uch term does not include telecommunication services" but the definition of "telecommunications" and "telecommunications service" in section 2(15) and (17), respectively, appear to include many Internet-related activities. The relationship of these definitions to one another may become more clear as the

Proposal is discussed further. But it may be appropriate for the Committees to consider refining these definitions.

**D. *Placing the Proposal within the Lanham Act***

The Proposal does not state where the statute would be placed. I believe that reflects the Committee's prudent recognition that a true misappropriation statute does not belong in the Copyright Act. *See, e.g.* Hughes, *How Extra-Copyright Protection of Databases can be Constitutional, supra* at 212.

**III. Is this Proposal "comparable" to the EU Database Directive?**

One of the reasons for continued U.S. attention to this issue is the European Union's creation of a *sui generis* form of database protection in 1996. At the time of its enactment, the Database Directive was justified largely on the grounds that Europe's database industries needed to become more competitive vis-à-vis the United States. Under Article 11 of the Directive, the *sui generis* protection is only granted to foreigners under agreements "concluded by the Council [of Ministers] acting on a proposal from the Commission." Recital 56 of the Directive further provides that this "reciprocal" will be offered to citizens of third countries "only if such third countries offer **comparable** protection . . . ."

Proponents of database protection in the United States point to these reciprocity provisions and the spectre that some U.S. database producer will have their database appropriated in Europe without having any protection under national laws implementing the Directive. To the best of my knowledge, there are no reported examples or cases of such database piracy in Europe. There are several reasons for this. First, most American companies that would be marketing databases in Europe maintain a sufficiently substantial presence to qualify for protection under the Directive. Indeed, many American databases are produced by U.S. companies that are wholly owned by European companies. Second, even if a U.S. database maker could not avail themselves of national laws directly implementing the Database Directive, many European countries have unfair competition laws that can be used against free-riders. In fact, in some early cases enforcing the Database Directive, national courts recognized that the plaintiffs could proceed under unfair competition laws as well. *See, e.g. Editorial Aranzadi v. Dealing World España, SA*, Court of First Instance, Alicante, Spain, Judgment of July 2, 1999.

Nonetheless, gaining "protection" for American database makers is still presented as an argument for legislating database protection in the United States. For that reason, the Proposal needs to be measured on a simple question: is the European Commission likely to judge the Proposal "comparable" to the Directive? The answer appears to be: no.

In late August 2003, at a symposium at the University of Oslo, one of the European Commission officials who drafted the Database Directive and oversees its implementation, Dr. Jens Gaster, made a presentation to a group of intellectual property professors and experts, principally from Europe, but including academics from Japan and the United States. Dr. Gaster's presentation was informative and included a discussion of the two cases of "comparable" protection that have been considered by the European Commission.

In the first, the Commission formally considered a request by the Isle of Man – a British Crown Dependency – for recognition of its "comparable" database protection. This was a very easy case of "comparable" protection because the Isle of Man had adopted the United Kingdom law implementing the Directive.[8] If the U.K. fulfilled the requirements of the Directive, the Isle of Man law did so automatically. According to Dr. Gester, the EU recognized the "comparable" protection in July 2003.

In contrast, Australia informally approached the Commission following the Australian Federal Court's decision in the *Telstra* litigation, a decision that concluded that Telstra had copyright in telephone books from labor and investment in gathering and compiling names, addresses, and telephone numbers.[9] The *Telstra* decision essentially ruled that Australian law follows English law prior to the Database Directive in providing "sweat of the brow" protection. According to Dr. Gester the Commission advised Australia that such protection was not "comparable" to the Directive on the grounds that the "[r]e-utilization right under Article 7 (2)(b) Directive goes beyond the traditional bundle of copyright rights" and "[c]opyright term is much longer than sui generis right's term of 15 years from completion of manufacture."[10] According to Dr. Gester, the informal "Australian request under Article 11(3) Directive [was] turned down."

Dr. Gester's clear and insightful presentation of the Commission's views show that the Proposal would be unlikely to be viewed by the Commission as offering "comparable" protection. Like the Australian law, the Proposal does not

provide the wide range of protection against "re-utilization" that the Directive does. Like the Australian law, the Proposal would provide a term of protection potentially much longer than the Directive's 15 years. On these grounds, the Proposal, in its present form, would not achieve one of the frequently mentioned goals of database protection legislation in the United States.

**Conclusion**

The issues surrounding extra-copyright protection of databases are complex. The Committee staff have made a laudable effort to open up new possibilities to craft a misappropriation law for the United States that can provide appropriate protection domestically and a model internationally to contrast with the strong protection provided by the European Union's Database Directive. There is much in the Proposal that can and should be improved, but there is much in the way of new ideas and approaches that merits serious study before it is condemned.

---

ENDNOTES

[1] Council Directive No. 96/9, O.J. L 77/20 (1996) (available on Europa, *Council Directive 96/9/EC on the Legal Protection of Databases* <http://europa.eu.int/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=EN&numdoc=31996L0009&model=guichett

[2] *See* Justin Hughes, *How extra-copyright protection of database can be constitutional*, 28 U. DAYTON L. REV. 159 (2003). For a sample of those strongly opposed to extra-copyright database protection on constitutional grounds, see Yochai Benkler, *Constitutional Bounds of Database Protection: The Role of Judicial Review in the Creation and Definition of Private Rights in Information*, 15 Berkeley Tech. L.J. 535, 575 (2000); William Patry, *The Enumerated Powers Doctrine and Intellectual Property*, 67 Geo. Wash. L. Rev. 359, 360 (1999); 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, 90 (1999). For analyses less strongly opposed on constitutional grounds, see Hughes, *supra.*; James Weinstein, *Database protection and the First* Amendment, 28 U. DAYTON L. REV. ___ (2003); Paul J. Heald & Suzanna Sherry, *Implied Limits on the Legislative Power: The Intellectual Property Clause as an Absolute Restraint on Congress*, 2000 Ill. L. Rev. 1119, 1197 (2000); Memorandum from William Michael Treanor, Deputy Assistant Attorney

General, Office of Legal Counsel, Department of Justice, to William P. Marshall, Assoc. White House Counsel, *Constitutional Concerns Raised by the Collections of Information Antipiracy Act, H.R. 2652* <http://www.acm.org/USACM/copyright/doj-hr2652-memo.html> (last updated July 28, 1998) ("OLC Memorandum").

[3] The §102(14)(B)(i) language says "a family and its social acquaintances," but the intent was surely not to say that a married person can distribute to the "family's" social acquaintances, but a single person cannot distribute to *her* social acquaintances. That would be problematic.

[4] Moreover, it is not clear that the narrow understanding of "time sensitive information" in *NBA v. Motorola* is the proper reading of the *INS* decision – or of the Court's acknowledgement of the *INS* misappropriation analysis in *Feist*. The Supreme Court in *INS* clearly stated that it was focused on "the question of unfair competition in business." 248 U.S. at 235. The *INS* Court was clear that it was making an unfair competition/misappropriation ruling completely separate from considerations of copyright law: "We need spend no time, however, upon the general question of property in news matter at common law, or the application of the copyright act, since it seems to us the case must turn upon the question of unfair competition in business. And, in our opinion, this does not depend upon any general right of property analogous to the common-law right of the proprietor of an unpublished work to prevent its publication without his consent; nor is it foreclosed by showing that the benefits of the copyright act have been waived." 248 U.S. at 234 – 235.

[5] H.R. Subcomm. on Telecomm., Trade, and Consumer Protection of the Commerce Comm., *Consumer and Investor Access to Information Act of 1999: Hearings on H.R. 1858*, 106th Cong. pt. IV A. (June 15, 1999) (Statement of Andrew J. Pincus, General Counsel, U.S. Department of Commerce) [hereinafter *1999 Administration Statement on H.R. 1858*] (available on USPTO, *Administrator for Legislative/International Affairs* <http://www.uspto.gov/web/offices/dcom/olia/hr1858.htm> (accessed September 22, 2003)).

[6] H.R. Subcomm. on Telecomm., Trade, and Consumer Protection of the Commerce Comm., *Consumer and Investor Access to Information Act of 1999: Hearings on H.R. 1858*, 106th Cong. pt. IV A. (June 15, 1999) (Statement of Andrew J. Pincus, General Counsel, U.S. Department of Commerce) [hereinafter *1999 Administration Statement on H.R. 1858*] (available on USPTO, *Administrator for Legislative/International Affairs* <http://www.uspto.gov/web/offices/dcom/olia/hr1858.htm> (accessed September 22, 2003)).

[7] To be sure, the Supreme Court has decided a number of cases where it has had to map out the relationship between different forms of protection. *See, e.g. Kellogg Co. v. National Biscuit Co.*, 305 U.S. 111, 59 S.Ct. 109 (1938) [patent and trademark]; *Mazer v. Stein, supra*; *Traff-ix Devices v. Marketing Displays*, 532 U.S. 23, 121 S.Ct. 1255 (2001) [patent and trademark]; *Dastar Corporation v. Twentieth-Century Fox*, ___ U.S. ___, 123 S. Ct. 2041; 156 L. Ed. 2d 18 (2003) [Lanham Act and copyright]. Overlapping protection from different intellectual property laws is an issue that needs to be watched very carefully. *See generally* 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 (1993).

---

[8] *See, e.g.* AGREEMENT IN THE FORM OF EXCHANGE OF LETTERS between the United Kingdom on behalf of the Isle of Man and the European Community extending to the Isle of Man the legal protection of databases as provided for in Chapter III of Directive 96/9/EC, March 26, 2003, Official Journal of the European Union, April 5, 2003, at L 89/12 – 15.

[9] Desktop Marketing Systems Pty Ltd v. Telstra Corp. Ltd., [2002] F.C.A.F.C. 112 (Federal Court of Australia, May 15, 2002) (holding that Telstra had copyright in telephone books from labor and investment in gathering and compiling names, addresses, and telephone numbers).

[10] Presentation of Dr. Jens Gaster, August 29, 2003, at the Columbanus Symposium, University of Oslo, Norway (co-sponsored by the University of Oslo and the University of Amsterdam). Powerpoint slide presentation on file with author.

October 29, 2003

The Honorable Lamar Smith, Chairman
Subcommittee on the Courts, the Internet and Intellectual Property
House Judiciary Committee
B-351A Rayburn House Office Building
Washington, D.C. 20515

The Honorable Cliff Stearns, Chairman
Subcommittee on Commerce, Trade and Consumer Protection
House Energy and Commerce Committee
2125 Rayburn House Office Building
Washington, DC 20515

Re: Submission for the record for the joint hearing before the Intellectual Property and Consumer Protection Subcommittees on the "Database and Collections of Information Misappropriations Act," held on September 23rd, 2003

Dear Mr. Chairman Smith and Chairman Stearns:

The Coalition Against Database Piracy (CADP) hereby requests that the two documents accompanying this cover letter be included in the record for the joint hearing before the Courts, the Internet and Intellectual Property Subcommittee and the Commerce, Trade and Consumer Protection Subcommittee on the *Database and Collections of Information Misappropriations Act*, held on September 23rd, 2003. The first document is a memorandum authored by C. Boyden Gray, Jamie Gorelick, Randolph D. Moss on October 27th. It establishes that the allegations made by the U.S. Chamber of Commerce in its letter dated October 15, 2003, are baseless by confirming that "numerous judicial decisions — along with common sense — demonstrate that the gap left by the *Feist* decision is both real and detrimental."

The second document is a report authored by Laura D'Andrea Tyson and Edward F. Sherry entitled *Statutory Protection for Databases: Economic and Public Policy Issues*. This report responds to the Chamber's claims that there is a lack of economic data to support the need for federal legislation to protect the investments of database producers. The Tyson report clearly demonstrates that is not the case as it provides numerous real-life examples and economic evidence validating the need for federal legislation to protect against database piracy.

Thank you for including these two documents in the record for the September 23rd hearing.

Sincerely,

Keith Kupferschmid
Counsel for CADP

WILMER, CUTLER & PICKERING
2445 M STREET, N.W.
WASHINGTON, DC 20037-1420
TELEPHONE +1 (202) 663 6000
FACSIMILE +1 (202) 663 6363
WWW.WILMER.COM

WASHINGTON
NEW YORK
BALTIMORE
NORTHERN VIRGINIA
LONDON
BRUSSELS
BERLIN

October 27, 2003

**MEMORANDUM ON THE LEGAL NEED FOR H.R. 3261, THE "*DATABASE AND COLLECTIONS OF INFORMATION MISAPPROPRIATION ACT*"**

**From:** **C. Boyden Gray, Jamie Gorelick, Randolph D. Moss**

**To:** **The Coalition Against Database Piracy (CADP)**

## EXECUTIVE SUMMARY

The distribution and analysis of data are at the core of the information economy. Before performing these functions, however, it is necessary to gather the underlying data — a process that often involves enormous effort and cost. H.R. 3261, the "Database and Collections of Information Misappropriation Act," seeks to protect the investment of those who develop the databases that form the foundation of the information economy. Without meaningful legislation, both businesses and consumers will suffer: businesses will risk the misappropriation of their investment by free-riding competitors, and consumers will be harmed by the decreased incentives that businesses will have to invest in developing and maintaining high-quality databases.

In the face of the simple proposition that individuals and companies should be allowed to protect their investment from misappropriation, opponents of the legislation have argued that H.R. 3261 is unnecessary (and potentially harmful), overbroad, and arguably exceeds Congress' authority. None of these arguments can withstand scrutiny.

First, some contend that there is no "gap" in existing law and that the creators of databases are already adequately protected from piracy. That contention is baseless. In *Feist Publications, Inc. v. Rural Tel. Serv. Co., Inc.*, 499 U.S. 340 (1991), the Supreme Court recognized that federal copyright law does not protect the most important element of a database, namely, the facts that are actually compiled. Numerous judicial decisions under copyright and other laws — along with common sense — demonstrate that the gap left by the *Feist* decision is both real and detrimental. The attempts by the opponents of H.R. 3261 to distinguish judicial decisions highlighting the gap in existing law all are unavailing.

Second, opponents of the legislation have argued that it is "not a misappropriation statute" and that some of its principal terms apply too broadly. Even a cursory review of the bill, however, confirms that it is, in fact, a misappropriation statute. The bill does not grant "exclusive rights," as was the case in some past proposals. The bill targets only those entities that would "free ride" on the investment, hard work and reputations of database producers and

cause them financial harm. Ironically, at the same time the opponents of the legislation argue that it is flawed because it does not fill a gap left by common law remedies, they argue that it should be narrowed in scope so that it merely replicates existing remedies.

Finally, while some have argued that because the Supreme Court recognized in *Feist* that the Copyright Clause of the Constitution includes an "originality" requirement, Congress lacks the authority to protect databases from piracy. The fact that Congress may lack authority to provide meaningful protection to databases under the Copyright Clause, however, says nothing about its authority to regulate under the Commerce Clause. And, given the enormous impact of database misappropriation on the national economy, there can be no doubt that H.R. 3261 fits easily within Congress' power to regulate interstate commerce.

## ANALYSIS

### I. AFTER *FEIST*, THE LAW PROVIDES THE CREATORS OF DATABASES WITH LITTLE IF ANY PRACTICAL PROTECTION

In *Feist*, the Supreme Court held that the *only* elements of a compilation of facts that are subject to protection under the Copyright Act are the selection and the arrangement of those facts — and these elements are subject to protection only to the extent they satisfy the requirement of "originality." Thus, even where a compilation of data reflects the "sweat of the brow" efforts of thousands of hours of work and highly-skilled data-collection, the Copyright Act does not prevent a database pirate from waiting for the information to be collected and then simply copying it in order to create a competing product. Those who do nothing more than compile facts and present them for consumption in an essentially non-creative way thus receive no federal protection for their investment. In fact, federal copyright law provides the least protection to the databases that are most commercially valuable: those that are comprehensive (reflecting little creativity in the "selection" of facts) and simple in their format (reflecting little creativity in the "arrangement" of the facts). *See Schoolhouse, Inc. v. Anderson*, 275 F.3d 726, 729 (8th Cir. 2002) ("a competitor may take 'the bulk of the factual material from a preexisting compilation' without infringing the author's copyright").

The case law confirms that current federal and state law do not provide adequate protection to the most commercially useful data compilations. Attempts by opponents of the legislation to distinguish these cases, moreover, are all unavailing.

○ In *Warren Publishing, Inc. v. Microdos Data Corp.*, 115 F.3d 1509 (11th Cir. 1997), Warren Publishing, which distributes an exhaustive directory of cable systems, could not prevent a competitor from incorporating its directory into the competitor's product. Under *Feist*, the facts themselves were not subject to copyright protection, only their selection or arrangement. *Id.* at 1515. Hamstrung by *Feist*, Warren tried to establish that, at a minimum, the defendant had infringed its system of sorting cable system data by "principal community," but that was not enough — even though, as the Eleventh Circuit noted, this method of sorting the data made it "commercially useful," it was not a sufficient basis for copyright protection. *Id.* at 1518. Although Warren went to enormous expense to create its "commercially useful" resource, a competitor could simply copy the facts that Warren had gathered and market them in its own products.

The opponents of H.R. 3261 argue that *Warren* fails to show a need for database protection legislation because, they assert, Warren subsequently began to use shrink-wrap licenses to protect its database from misappropriation, and, in a future case, might bring a claim for breach of contract. That answer, however, is wholly unsatisfactory. Contract remedies only bind the parties to the agreement; thus, for example, if a pirate obtains the protected database through a third party, the owner has no recourse against the pirate. Further, contract law varies from state to state — posing particular difficulties for those whose databases are broadly disseminated electronically, where the greatest potential for exploitation exists. In addition, the enforceability of click-wrap agreements — the principal means of establishing contracts in the context of electronic media — remains in flux. And, even if issues of enforceability could be overcome, a contract law solution is inefficient and will limit the distribution of databases: If contract law were the only source of protection, for example, then a database owner would have to bury its database behind such an agreement to ensure that the database would be protected from exploitation. A database protection law, on the other hand, would give the database owner additional incentive to make its database available more freely, and to a wider audience, without worrying about ensuring that every user of the database has explicitly assented to the terms of adequate protection. Indeed, if contract law is the only source of remedy, database providers would have strong incentives to offer their services only to large, institutional customers who are more likely to be held to the terms of their contract.

○ In *Schoolhouse, Inc. v. Anderson*, 275 F.3d 726 (8th Cir. 2002), the Eighth Circuit explained that a service that painstakingly compiled data on public schools did not have a copyright claim against a person who allegedly harvested data within 56 of those topics, comprising 74% of the topics listed in the database, to create his own database service. Because the defendant's service was not arranged in a similar fashion to the plaintiff's, and because the plaintiff's database was so comprehensive — thereby minimizing the "creativity" in the selection of topics — the Eighth Circuit held that the defendant's service was not infringing. Under copyright law, "when it comes to factual compilations, after *Feist*, it takes virtually 'extensive verbatim copying' to constitute infringement." *Id.* at 729.

Bewilderingly, opponents of H.R. 3261 argue that the *Schoolhouse* case does not support the need for legislation because, they assert, the district court found that there was no evidence that the defendant had copied the data from Schoolhouse, Inc. Yet, not only did the district court fail to make such a finding, the question was never before the court for one simple reason — as the court correctly recognized, under existing law "[c]opying of the underlying factual material contained in the compilation is not infringement." *Schoolhouse, Inc. v. Anderson*, 2001 WL 1640081, at *4 (D. Minn. Nov. 8, 2001). In short, Schoolhouse, Inc. was never given the opportunity to prove that its competitor copied substantial amounts of its data because, under existing copyright law, the competitor was allowed to do so. This is precisely the gap that H.R. 3261 is intended to fill.

○ In *Ticketmaster Corp. v. Tickets.com*, Copy. L. Rep. (CCH) P28,607 (C.D. Cal. 2003), a competitor of Ticketmaster used electronic means (called a "spider") to harvest in bulk ticket information from the Ticketmaster web site, for the purpose of displaying that information on its own online ticket service. Ticketmaster argued that its competitor had engaged in "trespass to

chattels" by electronically harvesting ticket data from its web site in this way. The district court expressly rejected economic harm (absent harm to the computer system) as a basis for trespass, on the grounds that a trespass to chattel claim required some actual "tangible interference with the use or operation of the computer being invaded by the spider," and Ticketmaster had failed to come up with evidence of such harm. In reaching this conclusion, the court noted that the trespass theory constituted an attempt "to apply a medieval common law concept in an entirely new situation which should be disposed of by modern law designed to protect intellectual property interests" — highlighting the fact that no such law exists.

In response to this case, the opponents of H.R. 3261 assert that *Ticketmaster Corp.*'s approach to the trespass to chattels claim is an outlier and that a recent decision of the California Supreme Court has called *Ticketmaster* into question. Unfortunately, just the opposite is true. The California Supreme Court actually refused to extend the tort of trespass to chattels to claims based on electronic "trespassing" (there, through the delivery of unwanted emails to Intel employees) that did not result in actual damage to the computers or otherwise harm their operation. *See Intel Corp. v. Hamidi*, 30 Cal. 4th 1342, 1364 (2003) ("We therefore decline to create an exception . . . to the general rule that a trespass to chattels is not actionable if it does not involve actual or threatened injury to the personal property or to the possessor's legally protected interest in the personal property."). The opponents of H.R. 3261 also suggest that, regardless of California law, Ticketmaster was protected by the Computer Fraud and Abuse Act ("CFAA"). But, the CFAA prevents only unauthorized access to protected computers. Thus, it would not have any bearing on databases and directories that are distributed by CD-ROM or in hard copy, or that are lawfully downloaded before they are pirated. Further, it remains unclear whether the CFAA applies to unauthorized harvesting of data from publicly available websites or other services.

○ In *Skinder-Strauss Associates v. Mass. Continuing Legal Education, Inc.*, 914 F. Supp. 665 (D. Mass. 1995), the district court concluded that a directory of attorneys and judges was not subject to protection under the Copyright Act — and therefore, freely reproducible in a competing publication — precisely because the directory was itself exhaustive. In that case, the plaintiff had assembled a directory of lawyers who practice in Massachusetts and judges who sit in that state, and presented evidence that the defendant had simply copied that information — including by copying fictitious names or "seeds" planted in the text to detect copying — into its own directory. Even assuming that the plaintiff's competitor had slavishly copied the directory information, however, the court held that such copying could not constitute copyright infringement: after *Feist*, database providers are free to engage in wholesale copying of facts from their competitor's databases into their own, even if to do so is essentially cost-free.

Opponents of the legislation note that the district court in that case held that other elements of the plaintiffs' publication were copyrightable, and that this demonstrates that current law provides adequate protection for database creators. That, of course, is an utter *non-sequitur*: the plaintiff enjoyed *no* protection in the very aspect of its publication that made it most useful, namely, the exhaustive directory of attorneys and judges.

In short, the attempt by opponents of the legislation to distinguish these cases are unavailing. And in any event, these cases are simply emblematic of a much broader problem,

evidenced by the raft of litigation in which database owners try (unsuccessfully) to prevent others from pirating their hard-earned compilations, each of which demonstrate the increasingly "thin" protection that the creators of databases enjoy as a matter of copyright law. For example, in *Bellsouth Advertising & Publishing Corp. v. Donnelley Information Publishing, Inc.*, 999 F.2d 1436 (11th Cir. 1993), the Eleventh Circuit held that the publisher of a "yellow pages" could not use copyright law to prevent a competitor from simply copying the name, address, telephone number, business type, and unit of advertising into its own database, even though that information was collected only as a result of a "number of marketing techniques" that the plaintiff deployed: because such "acts are not acts of authorship, but techniques for the discovery of facts," the Copyright Law provided the publisher no remedy at all. *Id.* at 1441. Similarly, in *Key Publications, Inc. v. Chinatown Today Pub. Enter., Inc.*, 945 F.2d 509, 515 (2d Cir. 1991), a competitor in the market for directories of Chinese-American businesses in New York evaded copyright liability, even though seventy-five percent of the competitor's directory was the result of "deliberate copying" of entries from the plaintiff's directory. The plaintiff in *EPM Communications, Inc. v. Notara, Inc.*, 2000 Copy. L. Rep. (CCH) P28,135 (S.D.N.Y. 2000), faced the same problem: the court held that the publisher of a massive directory of persons involved in the licensing business was not likely to prevail in its claim of copyright infringement against the publisher of a competing directory, even assuming that nearly half of the competitor's listings were copied from the plaintiff's directory. Likewise, the plaintiff in *American Massage Therapy Ass'n v. Maxwell Petersen Assocs., Inc.*, 209 F. Supp. 2d 941 (N.D. Ill. 2002), could not prevent its competitor from simply copying 17,617 names and addresses out of its directory of professional massage therapists and into its own database, to be sold to third parties for direct marketing purposes.

Opponents also suggest that there is a lack of "hard data" on the effect of this void in the law on the market for database products. To be sure, there exists a market for database services. But if H.R. 3261 were in place, the incentives of database providers to create and maintain databases of increasingly high quality would undeniably increase. Moreover, it is fundamentally unfair to deprive a person of fair compensation for the value of his services. In this context, the creators of databases, who expend vast resources to collect previously uncompiled facts, go uncompensated for the value they provide to third parties whenever someone copies their work and distributes it as their own. In a legal regime that authorizes such misappropriation, the incentive of database providers to create the products in the future or invest in their quality is diminished. For the same essential reasons, the protection afforded to database creators must be retroactive: both to ensure compensation for past efforts expended and as an incentive to create new databases and to maintain and to distribute existing databases going forward.

## II. THE PROPOSED LEGISLATION IS CLEARLY DESIGNED TO PREVENT A FORM OF "MISAPPROPRIATION"

The suggestion that H.R. 3261 is not a "misappropriation" bill is misplaced. Misappropriation is simply "[t]he application of another's property or money dishonestly to one's own use." Black's Law Dictionary (7th ed. 1999). And that is exactly what the bill sets out to prevent. Moreover, unlike copyright law, the legislation is specifically limited to circumstances in which piracy causes commercial harm, the database was generated at substantial cost, and the misappropriated database is made available in commerce "in a time sensitive manner." Plainly, H.R. 3261 is a misappropriation bill.

Ironically, at the same time that opponents of the bill argue that it does not fill any gap in existing law, they also argue that it unduly expands the common law of misappropriation. Yet, to the extent the legislation does so, it merely highlights the need for Congress to act. Thus, opponents of the legislation argue that H.R. 3261 would protect databases beyond the period in which the facts they contain are considered "hot news" under *International News Service v. Associated Press*, 248 U.S. 215 (1918). In fact, the legislation contains a time sensitivity provision, which requires courts to "consider the temporal value of the information in the database, within the context of the industry sector involved." It is hardly troubling that the bill, accordingly, protects a database during the period of time it is commercially valuable. Similarly, opponents find fault with the legislation because it would protect the investment of resources in the "maintenance" of a database as well as its creation. But a fundamental value that database providers offer their customers is a "maintained" database — one that is subject to ongoing revision for accuracy, ease of use, and exhaustiveness. It makes little sense to prevent database pirates from free riding on the initial creation of the database but to allow them to profit from similar efforts to maintain the database.

## III. CONGRESS HAS CONSTITUTIONAL AUTHORITY TO ENACT H.R. 3261

Congress has ample constitutional authority to fill this obvious gap. Constitutional scholar Paul Bender has written: "If a power given to Congress in order to serve a particular constitutional purpose does not authorize certain federal legislation, there is ordinarily no reason why a different power, designed to serve a different purpose, may not authorize the legislation instead." *See Paul Bender, The Constitutionality of Proposed Federal Database Protection Legislation,* 28 U. Dayton L. Rev. 143, 146 (2002). Even if Congress lacks authority under the Copyright Clause to protect the misappropriation of databases, it has broad authority under the Commerce Clause, U.S. Const. Art. I, § 8, cl. 3, to prohibit database pirates from misappropriating the labors expended by the database creators. For example, the Lanham Act — a source of extensive federal intellectual property rights — is based on the Commerce Clause, even though the Supreme Court specifically held that Congress could not base trademark protections on Copyright Clause. *See The Trade-Mark Cases*, 100 U.S. 82 (1879). Similarly, Congress has passed laws related to trade secrets (18 U.S.C. § 1831 *et seq.*) and the recording of live performances (18 U.S.C. § 2319A), neither of which find their basis in the Copyright Clause.

Some have suggested that this legislation would present substantial First Amendment problems, but any First Amendment concern is wildly exaggerated. The law of false advertising, trademarks, unfair competition, copyright, and trade secrets *all* involve regulation of information or words used in a commercial context, and all peacefully coexist with the First Amendment. And like these other established areas of commercial regulation, H.R. 3261 does not even apply to most uses of the databases that would raise First Amendment concerns: it includes substantial exemptions for non-profit, scientific, and educational uses, allows others to compile the same facts independently, and is otherwise narrowly tailored to apply only to prevent true database piracy that causes harm. Because current law provides only thin, inconsistent, and unpredictable protection for database creators, Congress reasonably may — and should — legislate in this important area of the law.

* * *

We would be pleased to answer any questions you might have.

# STATUTORY PROTECTION FOR DATABASES:

# ECONOMIC AND PUBLIC POLICY ISSUES

Laura D'Andrea Tyson and Edward F. Sherry[1]

## EXECUTIVE SUMMARY

### 1. Introduction

- In 1991, in a landmark case (Feist Publications, Inc. vs. Rural Telephone Service Co.), the Supreme Court ruled that copyright protection did not extend to all or parts of databases that did not involve some original "creative" selection and/or organization of data.[2] Indeed, the Court went further and ruled that such databases were not encompassed within the scope of the constitutional provision authorizing copyright protection. This sweeping decision eliminated the traditional "sweat of the brow" rationale for database protection that had been accorded under copyright law and left database producers in legal limbo in terms of their ability to protect themselves from unauthorized copying and dissemination of their products and from outright piracy.
- Both scholars and participants in the database industry agree that the current situation is undesirable.
- Because technology has expanded the potential applications of databases to myriad research, educational, medical, and business uses, the lack of adequate legal protections for the efforts of database providers poses a serious public policy challenge with widespread implications.
- This paper presents the economic rationale for statutory protection of databases, building on the general economic concepts of private property rights. It argues that databases produced and disseminated by private producers require legal protection to ensure that they are provided in amounts and forms consistent with their market demand. At the same time, there is a valid public interest both in maintaining access to information among the scientific, educational and library communities and in preventing the potential abuse of market power by private database providers. The public policy challenge is to find the appropriate legal means to balance the interests of database producers—who are concerned that without adequate legal protection they will not be able to earn an adequate return on the substantial costs of developing and maintaining their information products—and database users—who are concerned that statutory protection will impede the flow of information by restricting its availability and raising its price. In the end, both producers and users are seeking to ensure that there is information available to support education, scientific progress, and economic growth. An appropriately crafted law providing statutory protection can meet this challenge to the benefit of both producers and users.

---

[1] Dr. Laura D'Andrea Tyson is the Class of 1939 Professor of Economics and Business Administration at the University of California at Berkeley. Dr. Edward Sherry is an attorney and Senior Economist with the Law and Economics Consulting Group, Inc. Research assistance and support on this project was provided by Alan Marco.

[2] It is important to emphasize that, for databases, idiosyncratic or creative selection or organization of data may be undesirable. For many business and medical users, for example, the most valuable databases are those that contain comprehensive, current information that is logically organized so as to be easy to navigate.

## 2. Economic and Technological Issues

**Increasingly, the database market consists of electronic databases that allow users to combine software and information into powerful tools for research, educational, and commercial applications and for addressing major national challenges, such as finding cures for cancer and AIDS. Many of today's electronic products are developed along with sophisticated software for their use, and even when they rely on existing software packages, they are time-consuming and costly to produce and maintain because of rapid developments in software, computer and Internet technologies.**

- The database industry has grown rapidly, driven by new technologies. (Since 1979, the number of databases on the market has grown nearly 20-fold.)
- Some observers argue that rapid growth of the database industry shows that statutory protection is either unnecessary or can be delayed without significant cost. This argument fails to consider how fast the industry might have grown with statutory protection, and overlooks the fact that growth in the industry has been partly based on the expectation that policy makers will act to provide such protection in the future. Furthermore, this argument ignores the potential impact of the recent European Community Directive on future growth of the database industry in the United States.
- Electronic technologies have dramatically increased our ability to store, update and retrieve virtually limitless amounts of information. The paper describes several electronic databases that provide massive amounts of information about medicine, chemistry, agriculture and the military.
- Electronic technologies have also enhanced the ability of users to copy and sell databases, thereby increasing the vulnerability of database producers to piracy.
- At the same time, however, technology has enhanced the ability of database publishers to use "self-help" technical and contractual means to monitor the use and safeguard against the copying of their products. These means include encryption, passwords, on-line-only access to data, and contractual restrictions on further dissemination. In the absence of statutory protection, the industry has had to rely on such self-help approaches to protect their investments.
- Without statutory protection, database producers can be expected to underprovide their products in easily-copyable formats (such as CD-ROM). This has two effects: consumers are made worse off because they are deprived of database format choices; and industry growth is slowed by the resources spent on self-help means to prevent copying.
- The existence of technological and contractual means of protecting investments in the database industry does not obviate the need for statutory protection, just as fences or other methods of property protection do not obviate the need for legal protection against trespass—both work together.

### 2.1 The Changing Economics of Information Development

**The absence of adequate property protection and the threat of low rates of return for database products will reduce the supply of reliable information produced by private companies.**

From an economic point of view, all databases are *costly to produce* but *easy to reproduce or copy*. These two features of information generation and dissemination have some simple but profoundly important implications.

- First, competing in the ever-expanding market to meet the modern demand for information is expensive. A considerable amount of time, money and effort is required to construct and maintain a database—information must be generated and compiled, verified for accuracy, searched for errors, organized for use and interoperability with other hardware and software products, and continuously updated over time. Innovations in Internet and computer technologies have made these last two

features the most critical for today's users. Database providers must invest in their product to keep it functioning at state-of-the-art efficiency. Databases must be reconstructed to accommodate new conventions in computing power, new ways of "linking" to other electronic data (for example on the Web), and the exponential growth in the size of the data sets themselves.

All users—government, scientific, educational and commercial—have come to expect instant access to information. The Internet will continue to explode the stock of information available to all users at the relatively low cost of conveying it to them. But databases provide a different service. They organize, interpret and interrelate vast amounts of data according to an unlimited number of minutely tuned criteria that can be set and reset by the user. Information technologies, while expensive, have made possible research and discovery methods that were previously either unknown or impossible.

- Second, to protect our common interest in identifying, creating and making available the best information, we must protect this valuable resource from pirating. Revolutions in electronic technologies that have made databases easier to use and more potentially useful have also made them easier to "pirate." The ability of a potential competitor (or customer) to "free ride" on the substantial investment of an original database developer by copying and selling (or re-selling) his database weakens market incentives for investment in the database industry.

Broadly defined, intellectual property rights—including patent protection, copyright protection and to some extent trade secrets—seek to strike a balance between encouraging widespread use of information and encouraging its development by giving its developer a limited right to set the terms of its use in order to try to recoup his investment with a reasonable rate of return. It is this kind of balanced protection that the *Feist* case eliminated.

As an economic matter, whether the government or the private sector should produce a particular kind of information depends on which method of production is most efficient or least costly in terms of society's resources.[3]

- Sometimes government production will be the cheapest—for example, when the government is reporting on its own activities or when valuable information is a byproduct of government activities, such as law enforcement or space exploration.
- Even when the government is the least-cost provider of a kind of information, it is a mistake to conclude that such information is free or should be made freely available. In the absence of explicit fees or charges for the production and dissemination of data, the government must pay for the costs associated with these activities through taxation.
- For many kinds of information, however, there is a strong presumption that the private sector will be the low-cost provider. And there is also a strong presumption that private production and market signals will avoid wasting resources in generating information that has little or no value.
- Concern about the potential for the exercise of market power by private producers of databases is a major motivation behind concerns about statutory database protection, but there is little empirical evidence that the exercise of market power is actually a significant problem, even in so-called niche markets.

### 2.2 Responsive Pricing

**There is an incentive for private producers to try to extend the audience for their information products through a pricing strategy called price differentiation—that is, through charging different prices for the same or very similar information to different users or to users seeking different packages of information. However, database developers will**

[3] Concerns about possible government censorship or control over information also bear on the question of whether the government or the private sector should provide certain kinds of information.

**not be able to extend the audience for their products through price differentiation unless copying and resale are controlled through adequate statutory protection.**

Private producers of information typically will seek to recoup their investment costs by pricing their product above the cost of distributing that information to users. Such prices, while they may be necessary to insure continued investment and innovation in the database industry, will deter some potential users. Concern that such pricing strategies might limit access to information by scientists, educators, and libraries has motivated opposition to statutory database protection.

However, information providers already widely employ the practice of price differentiation, recognizing that it expands their potential market, not only for products of immediate concern, but also for future products, as more users become accustomed to incorporating database services into their work or research.

- Database users differ significantly not only in their ability to pay but also in their needs for accuracy, completeness, timeliness, support service, search engines, and ease of use. Database providers already offer different configurations of data and services at different prices, charging lower prices with fewer services to those with a low willingness or ability to pay and higher prices with more services to those with a higher willingness or ability to pay.
- Non-governmental producers of databases that are used for scientific research in both commercial and academic settings often charge lower prices for the latter.
- If copying and resale are not controlled by adequate statutory protection, they will undermine the ability of original database developers to use price differentiation. Both producers and consumers can benefit from price differentiation in the database industry, but piracy and copying will eliminate these benefits.

## 3. Concerns About Statutory Protection for Databases

There are strong economic arguments for statutory protection for databases. But concerns also have been voiced about possible negative effects of such protection. These concerns fall into two broad categories:

1. concerns that database producers, especially those of highly specialized goods with limited niche audiences, already have substantial market power, which will be enhanced by statutory protection to the detriment of consumers; and
2. concerns that such protection could reduce access to information by the scientific and educational communities, thereby slowing technological progress and economic growth.

### 3.1 Market Power

**Economists have long recognized that the credible threat of a new market entrant is a powerful constraint on the ability of firms to exercise their potential market power.**

Some skeptics of statutory protection for databases have concluded that "the market for commercially distributed databases is almost universally characterized by a distinct absence of competition" and that "the private database industry is largely characterized by niche marketers who supply and dominate specific market segments.[4]" These conclusions are not well supported by either economic logic or empirical evidence.

- A large part of the database industry, including many of the most commercially significant databases, operates in an intensely competitive environment.
- Even the existence of a small number of firms serving a particular market does not mean that those firms exercise significant market power.

[4] J.H. Reichman and Pamela Samuelson, "Intellectual Property Rights in Data?," Vanderbilt Law Review, Vol. 50, No. 1, January 1997.

1. Competing firms rarely supply the "same" database. Rather they compete on a range of fronts: selection and updating of data; convenience; search engine; ease of use; and price.

2. In a market economy, firms prosper by supplying what their customers want. When one sees a niche database market supplied by a single firm, that may be evidence that the firm is doing a good job serving the needs of the market at reasonable prices.

3. It is typically less costly for a new entrant to replicate an existing database than it was for its original producer to develop it in the first place, in which case the fixed costs of entry for the second firm are lower than they were for the first.

4. There are always opportunities for new entrants in niche markets as long as the underlying information contained in a particular database is available and can be replicated. If an investment in a particular database proves sufficiently attractive, firms will be encouraged to enter the market, and will be able to do so because they can turn to the original data or can enter into licensing agreements with the original compiler.

### 3.2 Competition and Data Replicability

**In most cases a potential entrant can get data from the same sources as the original firm, in which case there is no public policy need to allow the new entrant to free ride on the original firm's investment.**

The possibility of replicating the underlying data in a database is a key factor affecting the potential for market entry of new competitors. In fact, the underlying data in most databases is replicable.

- Data generated by the government is usually made available to users in its raw form at or below its dissemination cost, and that data can be collected by classic "sweat-of-the-brow" effort.
- Some databases rely on privately generated data that cannot be precisely replicated but for which comparable data would be available at comparable cost to competitors willing to make the necessary investment.[5]
- Privately generated data that relies on proprietary information cannot be replicated by competitors. As a pragmatic matter, it is unlikely that these databases would be commercially sold; information of this sort is commonly closely held as a trade secret.
- Historical data may not always be replicable. But if adequate protection is not provided for databases whose contents are not contemporaneously replicable, there will be little incentive to try to record the sorts of data that *cannot* be measured again—precisely the sorts of data that should be collected while it is still possible to do so.
- Some privately generated data cannot be replicated at comparable cost by a competitor because the data is produced by a publicly sanctioned monopolist. Phone numbers, for example, are arbitrarily assigned identifying data, privately generated with no additional effort (the phone company assigns a number as a provision of service) in the course of operating a publicly sanctioned local-monopoly business. When data is generated by a government-created monopolist, it is not appropriate to allow the monopolist to control database products building on that data. These sorts of data, collected by a government-created or government-sanctioned monopolist as a provision of service, should be made available to other users.
- Even when the underlying data in a database is not contemporaneously replicable, the basic case for adequate statutory protection remains the same: in the absence of such protection, there will be inadequate incentive to develop the original database in the first place.

[5] For example, the A.C. Nielsen Company collects scanner data through exclusive contracts with supermarkets. It then sells its ScanTrack reports about pricing, sales volume, and market share based on that data to grocery manufacturers. Nielsen's competitors can not *precisely* replicate Nielsen's data, but they can obtain "just as good" data through their own efforts.

### 3.3 Preferential Access

**The question of whether particular categories of users should get "preferential access" to information contained in privately produced databases is not the same question as whether the producers of these databases should be afforded adequate statutory protection.**

Access to data by the scientific and educational communities is vitally important, and effective statutory protection for databases can be drafted to respect these needs. Indeed, such protection will be beneficial to these communities, because it will provide the market incentives necessary to maintain healthy private investment in databases over time.

- Even if as a matter of public policy, certain kinds of users should receive preferential access, there remains the public policy issue of how that access is best achieved. A subsidy financing the purchase of a necessary good or service by preferred categories of users is the most direct form for realizing this objective. Allowing them to take the amount they want of a good or service from those who supply it without paying for it is only one particular form of subsidy—a subsidy in kind that is financed in essence by the suppliers and the other paying customers they serve. Usually such an in-kind subsidy proves considerably less efficient than a direct subsidy from the government.
- Those who use existing information at zero price can adversely affect the growth of information over time to both their own detriment and to the detriment of paying users.
- Inadequate funding can deprive such preferred users of the data they need to do their work, thereby depriving society of the benefits of technological progress and educational attainment. But in-kind subsidies taken from database providers by such users in the form of unauthorized copying for unauthorized purposes are not the appropriate remedy for addressing these valid concerns. To provide conditions for a healthy and competitive database industry that will serve the needs of all users, adequate statutory protection is required.
- The need for adequate funding for science and education—to support the use of equipment, publications, software and data—is an important public policy challenge. But this challenge is logically distinct from the challenge of providing an appropriate environment for protecting the rights of private database producers.
- Many of the concerns the scientific and educational communities have expressed in the debate about statutory database protection are in fact concerns about whether the government will continue to put adequate resources into the development and dissemination of those kinds of information for which it is the likely low-cost producer—information such as statistics, weather data, space exploration data, and court opinions—or whether it will "privatize" these activities in ways that will increase their cost to data users. Again, any debate over the privatization of information produced by the government is distinct from the debate about statutory protection for databases.

## 4. Legal Concerns about Appropriate Statutory Protection for Databases

Economic logic supports statutory protection for databases. How best can such protection be provided: How long should it last? What mode should it take? What should be its scope?

### 4.1 Duration of Protection

**The shorter the period of protection, the greater the incentive of producers to set high prices to try to recoup their investment during the allotted time.**

Under current US law, patents last 20 years from the date of filing; copyright protection lasts for the life of the author plus 50 years (or for 75 years from publication for works created by entities). The EC Database Directive protects databases for 15 years. HR 3531 proposed a 25-year term of protection.

- From an economic perspective, it is difficult to determine how long protection should last. Ideally, one might want the length of protection (and/or the scope of protection) to vary from database to database, but such a system would be totally impracticable.
- The shorter the period of protection, the greater the incentive of producers to try to set high prices to recoup their investment during the allotted time. Public policy makers should consider the likely relationship between duration of protection and firm pricing strategies in drafting statutory protection for databases.

### 4.2 Updating and Protection

**Database providers spend hundreds of millions of dollars a year updating their existing databases, and the newly-updated databases also need protection.**

- Some commentators have expressed concern that various proposals for statutory protection for databases will enable database providers to obtain "perpetual" protection for their databases merely by updating information contained in their products on a regular basis.
- This argument is no more significant when applied to databases than it is when applied to updated copyrighted material (such as new editions of books or reference works). Many encyclopedias or other reference works are updated regularly. Each new edition is copyrighted, and the copyrights on old editions expire overtime. This does not provide "perpetual" protection.
- If statutory protection for databases were provided for 25 years, the 1997 edition of a database would become available for copying by competitors in 2022. If in 2022 there were a market for the 1997 edition, a potential competitor could make a copy of it and compete with its original developer.
- Since the point of statutory protection is to protect investment in the creation, verification, maintenance and dissemination of information, such protection, when extended to updated products, in principle should apply to new content and to the additional investment required to verify, maintain and disseminate old content included in new editions. Once the protection on a database runs its course, users should be allowed to copy its contents.

### 4.3 Use vs. Replication: On Sufficient Statistics and Scientific Research

**The claim that increased protection will impede the ability to use databases for scientific and academic research rests on a misunderstanding of the difference between use and copying.**

Statutory proposals for database protection typically allow users to extract and copy small portions of a database, but seek to protect users or database developers against copying "all or substantial portions of" it. Some have suggested that this would improperly limit the ability of legitimate users to make use of the database. We disagree. *Using* a database is not the same thing as *copying* it.

- Scientists and other researchers typically can *use* "all or substantial portions of" the data in a database. They formulate and test hypotheses, perform statistical analyses, and, in the case of some sophisticated databases, even input their findings directly into the database to be compiled with and referenced against the complete data set.[6]
- Scientists and researchers may need to make and temporarily store (e.g., in RAM) an electronic "working copy" of the relevant data in a format useable by their statistics software package (or other analytic tool). Raw data is a tool, not a goal in itself; it needs to be used with other tools (such as search engines and statistics packages) in order to yield useful results. *Absent the ability to manipulate and analyze the data, scientists would have little use for it in the first place.* New legislation protecting databases can make it clear that

[6] For example, this interactive function is a component of MDL Information Systems' ISIS database software.

such temporary "working copies" are lawful, so long as they are not used by unauthorized persons or in an unauthorized manner.

### 4.4 Different Forms of Legal Protection For Databases

**Under current law, the main legal paradigm used to protect a database developer's interest in a database is a property-rule regime, as reflected in the common term "intellectual property" to refer to patent and copyright law. But there are also elements of a tort-based liability rule regime, notably in the law of trade secrets. One can use hybrid systems combining elements of tort and property.**

### 4.5 Liability Rules vs. Other Forms Of Protection

**The fundamental policy choice between a property rights approach and liability rules turns on whether negotiations among the parties are possible. When negotiations are not possible, liability rules have been developed to recompense owners for past infringement of their rights through court-awarded damages. When negotiations are possible, a potential user of property can negotiate with its owner about the terms and conditions for its future use.**

- Those contemplating copying or using someone else's database almost invariably know who developed the original database, and they are free to negotiate for the right to copy or use it.
- Arms'-length negotiations will establish a market price for those rights. We believe it is both unwise and inefficient to substitute court-established damages figures for market-established prices. It would establish the equivalent of a "private right of eminent domain," allowing others to take what they wish from a database without the owner's permission, subject only to paying court-awarded damages. The administrative burden imposed on the courts from trying to substitute court-awarded damages for privately negotiated prices would be severe.

### 4.6 Property Rights in Databases are Not "Exclusive" Rights to the Data

**Giving a database developer a "property" right in a database does not "exclude" others from replicating the underlying data from original sources and using or selling their own version of a database product based on that information.**

- In this regard, patent protection is significantly different from proposals for database protection. While both are termed "property" rights, the ability to preclude others from independent development of the same or a similar product are significantly different. Patent protection, in this sense, is much more exclusive in effect than are the proposals for database protection.

### 4.7 "Value-Added" Products

**There is a well-established system by which those who seek to provide "value-added" products negotiate with the original authors and publishers for the right to do so. The fees to be paid to the original authors and publishers are set by negotiation.**

- Many "value added products" are complements to the original database. The customer may find the database easier to use, or more valuable, because she also has the complementary product. Complementary products, such as an index to a database, a better search engine, or a manual, increase demand. Database developers have an incentive to encourage the development of complementary products.

- At the other extreme, some "value added products" are economic substitutes for the original database. This is especially likely to be the case if the substitute product contains a substantial amount of the contents of the original database. Consumers may or may not prefer the substitute product over the original database. Adequate statutory protection will allow the original developer to negotiate a license fee from a competitor for the right to develop a value-added product.

### 4.8 Compulsory Licensing

**A flat compulsory licensing fee would allow competitors to "skim the cream" by just copying the most successful databases. "Compulsory licensing" proposals are a particular form of price control, this time over the "price" of access to the contents of a database.**

Some critics of statutory database protection have urged that database providers be required to license their databases to "second comers" and those who wish to supply "value-added" products. For example, it has been proposed by Professors Reichman and Samuelson that others should be free to use all or part of a database on "payment of reasonable compensation according to a menu of user options vetted by the industry with user and government inputs."

- Implementing any such system, and making sure that the "reasonable compensation" and the "menu of user options" keep up with technological and market changes, is likely to be a daunting and controversial task.

- Economists have for years objected to the distortionary effects of price controls of all sorts. "Compulsory licensing" proposals are a particular form of price control, this time over the "price" of access to the contents of a database. The government is not well-placed to set such prices, for the same reasons that price controls are generally inefficient.

## 5. Conclusions

There are strong economic reasons for providing adequate statutory protection for the database industry. Such protection would serve the interests not only of database producers, but also of database users, including users in the scientific, educational and library communities. The creation, storage, verification, maintenance, updating and dissemination of information serve valuable economic and public policy functions—and they are not free whether they are performed by the government or private companies. Indeed, such activities often involve substantial upfront costs and considerable risk, since it is impossible to predict their actual value until the resulting information products are available for use. These costs and risks may be especially daunting for the development of highly specialized databases that are likely to have limited applications in the commercial arena, at least in the short run, and that therefore may have to rely initially on demand from a limited number of scientific and academic users with limited ability to pay.

Without effective statutory protection, private firms will be deterred from investing in database production. The resulting shortfall in the provision of information will have adverse effects on the pace of technological progress, on the economy's growth potential, and on the very research and educational communities whom critics of statutory protection wish to help.

# STATUTORY PROTECTION FOR DATABASES:

# ECONOMIC AND PUBLIC POLICY ISSUES[1]

Laura D'Andrea Tyson and Edward F. Sherry[2]

## 1. Introduction

In 1991, in a landmark case (Feist Publications, Inc. vs. Rural Telephone Service Co. [3]), the Supreme Court ruled that copyright protection did not extend to all or parts of databases that did not involve some original "creative" selection and/or organization of data.[4] Indeed, the Court went further and ruled that such databases were not encompassed within the scope of the constitutional provision authorizing copyright protection. This sweeping decision eliminated the traditional "sweat of the brow" rationale for database protection accorded under copyright law and widely relied upon by database producers, leaving them in legal limbo in terms of their ability to protect themselves from unauthorized copying and dissemination of their products.

Both scholars and participants in the database industry agree that the current situation is undesirable: in the words of Professors J.H. Reichman and Pamela Samuelson, two thoughtful but skeptical legal scholars who have written on database protection, "firms that make the contents of databases accessible to the public often become vulnerable to market-destructive appropriations that existing laws do not adequately remedy…The risk of market failure inherent in this state of chronic underprotection tends to keep the production of information goods at suboptimal levels.[5]" A similar conclusion has been posited by other legal scholars, such as Professor Paul Goldstein, who notes that

[1] Research for this project was funded by a contract between the Law and Economics Consulting Group, Inc., of Emeryville, CA, and Reed-Elsevier, Inc. of Newton, MA, and The Thomson Corporation of Stamford, CT, major international publishing companies. The views expressed in this article are those of the authors.

[2] Dr. Laura D'Andrea Tyson is the Class of 1939 Professor of Economics and Business Administration at the University of California at Berkeley. Dr. Edward Sherry is an attorney and Senior Economist with the Law and Economics Consulting Group, Inc. Research assistance and support on this project was provided by Alan Marco.

[3] 499 U.S. 340 (1991).

[4] It is important to emphasize that, for databases, idiosyncratic or creative selection or organization of data may be undesirable. For many business and medical users, for example, the most valuable databases are those that contain comprehensive, current information that is logically organized so as to be easy to navigate.

[5] J.H. Reichman and Pamela Samuelson, "Intellectual Property Rights in Data?," Vanderbilt Law Review, Vol. 50, No. 1, January 1997.

data and databases get less protection from copyright than their producers need to support the expense of data collection and assembly.[6]

As technology expands the potential applications of databases to myriad research, educational, medical, and business uses, the lack of adequate legal protections to database providers poses a serious public policy challenge with widespread implications. This challenge has been further complicated by the decision of the European Community (EC) to issue a directive that provides both copyright and specific protection for the contents of databases. The EC Directive demands reciprocity with a threat that US databases will not be afforded such protection in Europe unless the US adopts "similar" legislation by 1998. A first legislative proposal to address this challenge was introduced in Congress in 1996 (HR 3531), but it attracted criticism from segments of the scientific, educational, and library communities, as well as from some legal scholars. Recently, there has been renewed Congressional interest in providing some sort of statutory relief to database producers and to do so in a timely way to meet the needs of database producers and respond to the European Directive.

This paper presents the economic rationale for statutory protection of databases, building on the general economic concepts of private property rights. It argues that databases produced and disseminated by private producers require legal protection to ensure that they are provided in amounts and forms consistent with their market demand. At the same time, there is a valid public interest both in maintaining access to information among the scientific and educational communities and in preventing the potential abuse of market power by private database providers. The public policy challenge is to find the appropriate legal means to balance the interests of database producers—who are concerned that without adequate legal protection they will not be able to justify incurring the substantial costs of developing and maintaining their information products—and database users—who are concerned that statutory protection will impede the flow of information by restricting its availability and raising its price. In the end, both producers and users are seeking to ensure that there is information available to support education, scientific progress, and economic growth. The final sections of this paper present some suggestions about the kinds of legal protection that might provide an appropriate balance of interests between users and producers to the benefit of both.

## 2. The Changing Economics of the Database Industry

**Increasingly, the database market consists of electronic databases that allow users to combine software and information into powerful tools for**

[6] Paul Goldstein, Copyright's Highway (1994), p. 211.

**research, educational, and commercial applications and for addressing major national challenges. Many of today's electronic products are developed along with sophisticated software for their use, and even when they rely on existing software packages, they are time-consuming and costly to produce and maintain because of rapid developments in software, computer and Internet technologies.**

Although the actual wording of effective statutory protection for databases will require a formal definition of a database, for the purpose of understanding the economic rationale for such protection, a database can be defined simply as any organized collection of information. This definition is broad enough to include the vast array of databases that are part of today's economy, from text-based databases, such as databases of court opinions and statutory regulations, to numerical databases, such as databases of federal, state and local economic statistics, to sophisticated digital databases that combine information and software into complex systems for decision making. A database can be as simple as an index, an almanac or the listing of daily stock market prices at the back of the financial section of a newspaper. But, increasingly, the database market consists of electronic databases that allow users to combine software and information into powerful tools for research, educational, and commercial applications and for addressing major national challenges, such as finding cures for cancer and AIDS and maintaining the competitiveness of US agriculture.

Competing in the ever-expanding market to meet the modern demand for information has become expensive. A considerable amount of time, money and effort is required to construct and maintain a database—the need for the information must be identified; it must be generated and compiled, verified for accuracy, searched for errors, organized for use and interoperability with other hardware and software products, and continuously updated over time. Innovations in Internet and computer technologies have made these last two features critical for today's users. Database providers must vigorously invest in their products to keep them functioning at state-of-the-art efficiency. Databases must be reconstructed to accommodate new conventions in computing power, new ways of "linking" to other electronic data (for example on the Web), and the exponential growth in the size of the data sets themselves.

All users—government, scientific, educational and commercial—have come to expect instant access to relevant information. The Internet will continue to explode the stock of information available to all users at the relatively low cost of conveying it to them. But databases provide a different service. They organize, interpret and interrelate vast amounts of data according to minutely tuned criteria that can be set and reset by the user. Information technologies, while expensive, have made possible research and discovery methods that were previously either unknown or impossible.

A few examples illustrate the scope and complexity of today's electronic databases and the amazing applications they enable. PoisIndex is an index of approximately one million entries on a wide variety of poisonous substances, including drugs, chemicals, commercial and household products, and biologic substances. Substances are reviewed for entry into the database by a group of skilled medical professionals, who also scan the world's medical literature for pertinent data on toxic exposure and management. Approximately 200 actively practicing clinicians from over 20 countries participate in the editorial and selection process. Each substance entry in the database is linked with up to four full-text documents outlining clinical effects, range of toxicity, treatment measures, and other toxicological information. Software engineers are employed to maintain, test, produce and support the database and the software required to store, edit, sort and retrieve the data. The typical PoisIndex user is a medical professional, usually an emergency physician or poison center specialist, who needs instant access to such information in life-threatening circumstances.

The MDL Drug Data Report (MDDR), produced by MDL Information Systems in cooperation with Prous Science Publishers, is a database of approximately 85,000 chemical compounds with potential drug applications. It is updated on a monthly basis from a specialized search of published reports, patent applications and scientific papers so as to make data available on new biologically active compounds as soon as they are disclosed. MDDR tracks these compounds through stages of development and into clinical trials. Accompanying software permits researchers to analyze the effects of modifications of a drug compound's structure on its properties. Researchers can also combine the results of their own internal and external results with the database supplied by MDL to develop their own specialized research tool.

Visible Human is a database product consisting of more than 10,000 images for exploring human anatomy. Color photographs, along with magnetic resonance and registered computed tomography images, provide new perspectives into the structure and function of the body. Several versions of Visible Human are available for use with different hardware and software support systems including Windows, Macintosh, and Unix. Relying on any one of these systems, the user can view the body using a graphical navigator, reference images using bookmarks, animate a series of images to gain insight into anatomical relationships, and annotate and highlight areas of interest with text, color and markers. This research and educational tool builds on a database of the human body developed by the National Library of Medicine. Until a private producer came along to develop Visible Human as a commercial venture, the underlying data was not readily available for use by either academic or other potential customers.

Derwent World Patents Index is a comprehensive database of more than seven million separate inventions culled from more than 13 million patent documents worldwide. Coverage includes patents of products from the pharmaceutical industry, agricultural and veterinary medicine, polymers and plastics, chemistry, electronics, electrical and

mechanical engineering. All patent information is presented in a uniform, user-friendly format consisting of a simplified English-language abstract explaining key technical details and highlighting applications. In addition to bibliographic information, technical drawings or diagrams are included as available. The Index is updated weekly with information from 40 patent-issuing authorities around the world, 1200 scientific journals, and papers presented at international conferences. Users of the Derwent Index include patent and information professionals, research scientists, engineers, universities, research institutes, libraries, and individual inventors and entrepreneurs.

Now in its centennial year, Jane's provides a variety of data on defense and security issues and in the civilian aerospace and transportation fields. Jane's is rapidly completing a transition from print forms of database information, such as yearbooks and reference books, to electronic databases available to users in various formats including hard copies, online service, and CD-ROM. Jane's data collection process involves a large number of correspondents, freelance contributors, authors and editors who seek out or contribute information from a variety of sources including text, photos, video, audiotapes, and interviews. Jane's databases often rely on highly specialized data collection efforts. For example, to prepare a 200-page special report on trends in land mine warfare, one researcher traveled to Bosnia to view the latest discoveries in anti-personnel and anti-tank mines. In addition to its information-gathering staff, Jane's employs a pool of experts to verify the accuracy of its database products. Users of Jane's database products include government officials, military experts, journalists, industry experts, and academics.

Electronic databases are also being developed and used to improve agricultural decision making. The National Agricultural Database Laboratory, run by the University of Wisconsin and supported by a variety of public and private funding sources, has a number of database products either already available or under development for the nation's farmers. These include a National Dairy Database, a National Pig Database, a National Sheep Database, a National Beef Database, and a National Poultry Database. The National Dairy Database, the first of the Laboratory's projects, consists of 600 peer-reviewed US dairy publications and related publications on water quality, crops, fertilizers, and waste management. The solicitation process for this database began as a call for materials to agricultural extension offices, although entries from a variety of channels, including end users, now have been suggested for inclusion. The database is developed and supported by existing hardware and software technologies and maintained, updated, and enlarged by a university-based research team. Revenues from the sales of the database are returned to the Laboratory to cover expenses incurred in the preparation and electronic publishing of its growing number of database products. Users include individual farmers, university research institutes, agricultural extension centers, and individual researchers.

Database software is enhancing existing databases with powerful, specialized new tools, and database software developers are expanding into the provision of data to optimize

on these software capabilities for their users. MDL Information Systems, for example, found that there was an inadequate supply of scientific data for its ISIS software system, and has since developed eight different databases in reaction chemistry or synthetic methodology; two databases in bioactivity (including the Drug Data Report); one metabolism database; four chemical sourcing databases; and six databases of chemical safety information (precautions, toxicity, transportation warning and the like). The ISIS databases enable sophisticated scientific analysis by allowing users to create original data sets to test scientific hypotheses and methods in unique ways. ISIS software can combine archival data from one or more of the MDL databases, from another database provider, or from the scientist's own proprietary database into huge interrelated data sets that can be searched and linked in numerous combinations.

MDL's database of synthetic methods, for example, brings together for scientists a forty-year survey of published papers on synthetic methods with the most current methods, and this information is updated four times per year. Scientists can explore the effects of a particular synthetic method or set of methods on particular elements in ways that would previously have required months if not years of study of the literature and tedious experimentation in the laboratory.

Combinatorial chemistry databases perform automated syntheses of chemical elements that allow researchers to experiment with molecular structure. With this process they can develop new compounds, which may prove to have value as herbicides, drugs, polymers or fibers. Major pharmaceutical companies and biotechnology companies use the ISIS system to increase the speed of the product development process and reduce their time to market with new products.

As these examples indicate, today's electronic database products are a far cry from the simple printed databases that have been a significant part of the publishing industry for centuries. With the touch of a few keys, users can scan through information culled from a variety of sources over long periods of time. Many of today's electronic products are developed along with sophisticated software for their use, and even when they rely on existing software packages, they are time-consuming and costly to produce and maintain because of the necessity of keeping up with rapid developments in software, computer and Internet technologies. Indeed, sophisticated database producers such as MDL and Bloomberg maintain their own staff of professional software engineers. As described above, many of today's database products also have interactive features, allowing users to add their own data to the system's underlying information to create a customized database of their own. In addition, many database products are highly specialized for particular users. Unlike many software developers, who may develop a program with millions of potential customers, many producers of specialized databases face markets of limited size—sometimes no more than a few hundred customers. Without adequate legal protection, commercial producers of such products may not be able to justify the costs and risks associated with developing them.

Because the boundaries between databases and other sorts of information sources are increasingly fuzzy, it is difficult to get a precise measurement of the size of "the" database industry. But by any reasonable measure, it is both large and growing rapidly (see Tables 1 and 2).

Table 1 lists sales figures that represent many ways of measuring the role of databases in the economy. Some of the industry categories in the table clearly include more than just databases. For instance, the first entry "US publishing industry and related services," includes newspaper publishing, which is broader than just the databases contained therein. On the other hand, no category clearly includes all of the database industry. For instance, proprietary databases, such as those compiled by A. C. Nielsen Company, would not be included in the publishing category. Rather, Nielsen shows up as the biggest company listed by Ward's Business Directory under "Commercial Nonphysical Research," (our last entry in Table 1) with almost $1.3 billion in sales. Generally, most of the large companies in this category produce databases as one of their lines of business. However, this category would exclude companies such as LexisNexis, a large producer of full-text legal and news databases.

**Table 1. Estimates of the size of the database industry[7]**

| Industry Description | Year of Source | Sales | Other Data |
|---|---|---|---|
| 1. US publishing industry and related services | 1996 | $200 billion.[8] | |
| 2. Newspapers, books and magazines | 1996 | $85 billion.[9] | |
| 3. Data processing and network services, 1993 | 1994 | $46.4 billion.[10] | |
| 4. Business information suppliers | 1996 | $26 billion.[11] | |
| 5. Data processing and preparation, SIC 7374 | 1997 | $21.4 billion.[12] | 167 companies 180,700 employees. |
| 6. Electronic Information industry | 1996 | $15 billion.[13] | |
| 7. Database revenues of business information, 1994 | 1996 | $13.8 billion.[14] | |
| 8. 1993 Electronic information services | 1994 | $13.6 billion.[15] | |
| 9. 1995 Electronic delivery of business information (primarily online and CD-ROM) | 1996 | $10.7 billion.[16] | |
| 10. Information retrieval services, SIC 7375 | 1997 | $7.8 billion.[17] | 345 companies 60,800 employees |
| 11. Commercial Nonphysical Research, SIC 8732 | 1997 | $4.5 billion[18] | 413 companies 52,000 employees. |

Table 2 reveals that although both the number of databases and the number of database producers have continued to expand since the *Feist* decision in 1991, the growth rates for both of these measures slowed considerably in the six years following

[7] The categories in Table 1 give estimates from $4.5 billion to $200 billion in sales. Given that A. C. Nielsen accounts for $1.3 billion on its own, it is clear that the database industry is a multi-billion dollar industry. Additionally, within the "Business Information" market, the largest segment was "Database Publishing," category 7 in Table 1, which had revenues of $13.8 billion dollars in 1994. It is obvious that databases in all markets must amount to more than this figure, though it is difficult to know what is included in any given category without a detailed analysis of the raw data.

[8] Information Today, January, 1996.

[9] Sloan Management Review, March 1996.

[10] US Industrial Outlook 1994, p. 25-1.

[11] Sloan Management Review, March 1996.

[12] Ward's Business Directory, 1997.

[13] Information Today, January, 1996.

[14] Electronic Information Report, 1/5/96.

[15] US Industrial Outlook, 1994, p. 25-1.

[16] Interactive Daily, 2/5/96.

[17] Ward's Business Directory, 1997.

[18] Ward's Business Directory, 1997.

that decision compared to the prior six years. Although not conclusive, these growth numbers suggest that this decision may have dampened investment in the industry, as economic logic predicts.[19]

**Table 2. Growth of the database industry[20]**

| | Number of Databases | Number of Producers | Number of Online Services |
|---|---|---|---|
| 1979 | 300 | 221 | 59 |
| 1980 | 411 | 289 | 71 |
| 1981 | 641 | 411 | 135 |
| 1982 | 919 | 812 | 189 |
| 1983 | 1360 | 820 | 244 |
| 1984 | 1807 | 1069 | 327 |
| 1985 | 2247 | 1316 | 414 |
| 1986 | 2369 | 1379 | 454 |
| 1987 | 2823 | 1568 | 528 |
| 1988 | 3135 | 1685 | 555 |
| 1989 | 3535 | 1813 | 600 |
| 1990 | 3943 | 1950 | 645 |
| 1991 | 4332 | 2120 | 718 |
| 1992 | 4447 | 2033 | 772 |
| 1993 | 5183 | 2204 | 818 |
| 1994 | 5300 | 2232 | 822 |
| 1995 | 5342 | 2202 | 828 |
| 1996 | 5511 | 2255 | 860 |
| 1997 | 5739 | 2312 | 899 |

Source: Gale Directory of Databases, p. x. Information reflects the number of entries published in the Directory of Online Databases since 1979.

Electronic technologies have dramatically increased the ability to store, update and retrieve large amounts of information. They have also changed the ability of users to copy and sell databases and the ability of database producers to monitor their use. All

[19] Some observers argue that rapid growth of the database industry shows that statutory protection is either unnecessary or can be delayed without significant cost. But this argument fails to consider how fast the industry might have grown with statutory protection or the costs of various self-help technological means the industry has used to protect itself. It also overlooks the fact that growth in the industry has been partly based on the expectation that policy makers will act to provide such protection in the future. Furthermore, this argument ignores the potential impact of the EC Directive on future growth of the database industry in the United States.

[20] Because the cost of setting up and operating an online service is significant, it is to be expected that owners of on-line services will leverage that investment by carrying multiple databases, that the owner of a few databases would distribute them through another's on-line service and hence that the number of databases on line would grow faster than the number of on-line services.

of these changes have important implications for the economics of the database industry.

When databases were provided in book form, copying was limited by copyright law and the technology of printing, and more recently by the technology of photocopying. But with the advent of computerized databases, online access, and scanning technology, it is now significantly easier for a user to copy large parts of an electronic database.[21]

On the other hand, technology also has enhanced the ability of database publishers to use "self-help" technical or contractual means to monitot the use and safeguard against the copying of their products. These means include encryption, passwords, on-line-only access to data, and contractual restrictions on further dissemination. In the absence of statutory protection, the industry has had to rely on such self-help approaches to protect its investments. Although they may have been reasonably effective in some cases—and may explain in part why the industry so far has grown rapidly in the absence of statutory protection—such approaches may well have come at a cost in terms of efficiency. For example, many databases currently available only on-line could easily be disseminated to users on CD-ROM, and many users might prefer that format. But in the absence of statutory protection, database providers may not produce CD-ROM versions of their products because they can be easily copied and sold at a lower price. By contrast, it is considerably more difficult for potential pirates to copy and sell the contents of on-line-only databases. In other words, without statutory protection, database producers can be expected to underprovide their products in certain easily-copyable formats. This has two effects: consumers are made worse off because they are deprived of database format choices; and industry growth is slowed by the resources spent on self-help means to prevent copying.

In the database industry, as in other economic activities, self-help means of protecting property sometimes serve as economic substitutes for legal means of protection. For example, fences, guard dogs and private security patrols can substitute for statutory protections against trespassing on private property. Economic logic suggests, however, that there is an optimal mixture of self-help and legal forms of protection. In the case of property protection, self-help methods do not obviate the need for legal protection against trespass—both work together. Nor in the case of databases does the existence of technological and contractual means of protecting investments obviate the need for statutory protection. Indeed, rapid changes in digital technology can render many such

[21] It is also feasible to transfer a paper database into electronic format either by keypunching the data or by using optical character recognition software to scan a paper database and transfer it into electronic form. However, the former method is often very time consuming, while the latter technology is not perfected. It is likely to produce a significant number of errors in any lengthy work. But digital copying of electronic databases can be "perfect" and very low cost. A key feature of digital technology is that a copy is indistinguishable from the original, so that an Nth-generation copy is "just as good" as a first-generation copy. This is not the case for traditional analog photocopying, in which quality degrades relatively quickly after two or three generations.

self-help approaches obsolete overnight, as the history of copy-protection in the software industry demonstrates. They may also be incompatible with user needs and desires.

### 3. The Economics of Information and Databases

**The absence of adequate property protection and the threat of low rates of return for database products will reduce the supply of information produced by private companies.**

From an economic point of view, all kinds of databases, like any form of economically significant information, have two things in common: they are *costly to produce*, but they are *easy to reproduce or copy*. These two features of information generation and dissemination have some simple but profoundly important implications.

The fact that it is cheap to reproduce information once it has been produced suggests that it should be made available to users at its relatively low cost of dissemination, a cost that has fallen over time for most kinds of information as a result of revolutions in copying and digital technologies. Indeed, if there was a fixed supply of information available for free, economic theory would suggest that anyone who wanted it should be able to obtain it at no cost beyond the relatively low cost involved in conveying it to them (a cost that economists call the "marginal cost of dissemination").

But there is not a fixed free supply of information for two reasons. First, information must be developed, produced, generated or discovered in the first place. Second, it must be collected and made available to the consumer. Just as there is a significant difference in economic value between an apple hanging on a tree in an orchard 500 miles from me and the "same" apple in the produce section of my local supermarket, there is a significant "value-added" from compiling available information (often from diverse sources) and converting it into a useable form.

A great deal of time, money and effort is required to generate and compile information, verify its accuracy, detect technical or transcription errors, and organize it for use. The substantial costs involved (both monetary and non-monetary) are both "fixed" (in the sense that they do not depend on whether one person or a million uses the resulting information) and what economists call "sunk" (in the sense that these costs are not recoverable should the information no longer be needed).[22]

[22] The distinction between fixed and sunk costs can be seen by an example. Suppose that one wants to enter the sidewalk hot-dog vending business. To do so, one needs a business license and a pushcart. The costs of these are "fixed" in the sense that they do not vary with sales. ("Variable" costs, for example, for supplies, are of course subject to sales.) If one chooses to exit the business, it may be possible to sell the cart and thus recoup part of the

Without the prospect of an adequate return, producers will tend not to invest in generating, collecting and organizing information. This is especially problematic if end-users can take information without paying for it, or if competitors (or customers) "free ride" on the substantial investment of an original database developer by copying and selling (or re-selling) his database. The competitors, who do not have a comparable investment to try to recoup, can typically undercut the price, further undermining the original compiler.

From an economic perspective, then, there is a need to provide those who make such investments with the prospect of earning an adequate or competitive return on those investments, so as to encourage future efforts.[23] This presents the fundamental economic paradox of information. Once information has been generated, economic logic suggests that it should be made available to anyone who wants it at a low or zero price commensurate with the cost of disseminating it. At the same time, however, such a price does not provide adequate incentive for making investments in information products, nor does it provide a signal as to *which* databases should be produced in the future.[24]

Broadly defined, intellectual property rights—including patent protection, copyright protection and to some extent trade secrets—seek to strike a balance between encouraging widespread use of information and encouraging its development by giving its developer a limited right to set the terms of its use in order to try to recoup his investment with a reasonable rate of return.[25] It is this kind of balanced protection that the *Feist* case eliminated.

The fact that information is costly to produce raises another key question: who should produce it in the first place? Because information has a key characteristic of what economists call a "public good"—that is, one person using it does not prevent another person from using it—it is sometimes argued that the government should produce

---

fixed costs. The "sunk" costs are the part of the fixed costs that cannot be recouped--in this case, the cost of the license plus the difference between what it cost to buy the cart and the price for which it can be sold.

[23] The classic reference is Kenneth Arrow, "Economic Welfare and the Allocation of Resources for Invention," in Universities-National Bureau of Economic Research Conference Series, The Rate and Direction of Economic Activity: Economic and Social Factors (1962).

[24] Professor Kenneth Arrow also pointed out another problem with information. A potential buyer of information does not know what the information is, and therefore does not know what the information is worth to him, until he sees its contents. But once the information is revealed, the buyer has no incentive to pay for the information he has already seen, or (at the least) has an incentive to underreport the value he places on the information. Arrow made another point. Those who develop information tend not to be able to appropriate the full social value of that information. Consequently, there is (in general) a tendency to under-invest in the production and dissemination of information (relative to the level that would be first-best from a societal standpoint.)

[25] Note that we use the expression "try to recoup" because investment in information development is a gamble like other investment. There is no guarantee that demand will prove sufficient to enable the developer to earn a profit on his investment.

information or at the very least subsidize its production by the private sector.[26] But this conclusion is not always warranted. As an economic question, whether the government or the private sector should produce a particular kind of information depends on which method of production is most efficient or least costly in terms of society's resources. Sometimes government production will be the cheapest—for example, when the government is reporting on its own activities or when valuable information is a byproduct of government activities, such as law enforcement or space exploration.

But even when the government is the least-cost provider of a particular kind of information, it is a mistake to conclude that such information is free or should be made freely available. In the absence of explicit fees or charges for the production and dissemination of data, the government must pay for the costs associated with these activities through taxation. The costs of information do not disappear when the government is the provider, although how these costs are recouped and from whom may change significantly. Indeed, economic logic suggests that the government should charge at least the incremental cost of dissemination of government information to those who use it. The imposition of such user fees would reduce the need to finance information generation and dissemination through taxes, and would insure that those who use the information value it at least as much as it costs to provide it to them.[27]

For many kinds of information, however, there is a strong presumption that the private sector will be the low-cost provider. And there is also a strong presumption that private production and market signals will avoid wasting resources in generating information that has little or no value.

In the case of information production, the presumption in favor of private production must be tempered by the concern that the economics of information production—its high upfront costs and low marginal costs—may deter market entry and result in market power. Concern about the potential for the exercise of market power by private producers of databases is a major motivation behind concerns about statutory database protection. Later sections in this paper conclude, however, that there is little empirical evidence that the exercise of market power is actually a significant problem, even in so-called niche markets.

Even without market power, private producers of information will typically seek to recoup the upfront costs of their investment by pricing above the incremental cost of distributing that information to users. And such prices, while they may be necessary to secure an adequate return on investment, will deter some potential users of information and restrict information flows. Concern that such pricing strategies might limit access

---

[26] Concerns about possible government censorship or control over information also bear on the question of whether the government or the private sector should provide certain kinds of information.

[27] Carl Shapiro and Hal R. Varian, "US Government Information Policy," draft prepared for presentation at Highlands Forum, Department of Defense, Washington, DC, June 8, 1997.

to information by scientists, educators, and libraries has also motivated opposition to statutory database protection.

Luckily, however, there is an incentive for private producers to try to extend the audience for their information products through a pricing strategy called price differentiation—that is, through charging different prices for the same or very similar information to different users or to users seeking different packages of information products. Indeed, price differentiation is a traditional strategy for most kinds of information providers (as it is for producers of other kinds of goods whose production involves relatively high upfront costs and relatively low incremental costs) to attempt to recover their upfront costs by increasing the demand for their output. For example, the movie industry sells movies to theaters, hotels, airlines, video rental stores, and individual consumers at different prices. Similarly, the book industry sells hardback and paperback books at different prices, and the software industry sells its products at both retail and site-licensed prices.[28]

A similar strategy of price differentiation is also apparent in the database industry. For example, online information services, such as LexisNexis, often charge one price for daytime use and another lower price for nighttime use. Non-governmental producers of databases that are used for scientific research in both commercial and academic settings often charge lower prices for the latter. [29]

Since database users often differ significantly not only in their ability to pay but also in their needs for accuracy, completeness, timeliness, support service, search engines, and ease of use, database providers often have many opportunities to sell slightly different configurations of data and services at different prices, charging lower prices with fewer services to those with a low willingness or ability to pay and higher prices with more services to those with a higher willingness or ability to pay. Indeed, such pricing strategies may be essential for many database producers to generate enough demand and revenues from sales to cover their upfront production costs.

Such strategies can work, however, only as long as those who purchase a database at a relatively low price cannot copy and sell it to those who would otherwise have paid its original developer a relatively high price. If copying and resale are not controlled by adequate statutory protection, they will undermine the ability of original database developers to use price differentiation as a way to recoup upfront costs. And without recouping costs, database products will not survive in the market, harming both those who would have been willing to pay a relatively high price and those who would have been willing to pay a lower price. In other words, both producers and consumers can

[28] Shapiro and Varian, *ibid.*

[29] As an illustration, one successful database provider specializing in medical information offers many of its products to academic users at prices that are only about 1% to 2% of the prices paid by commercial users. The academic price is based on the approximate cost of its distribution to academic users (which is similar to its marginal cost of dissemination).

benefit from price differentiation in the database industry, but piracy and copying will eliminate these benefits.

### 4. Concerns About Statutory Protection for Databases

As the preceding discussion indicates, there are strong economic arguments for statutory protection for databases. But concerns also have been voiced about possible negative effects of such protection by some legal scholars and by some in the scientific, educational and library communities. These concerns fall into two broad categories: first, concerns that database producers, especially those of highly specialized goods with limited niche audiences, already have substantial market power, which will be enhanced by statutory protection to the detriment of consumers; and, second, concerns that such protection could reduce access to information by the scientific and educational communities, thereby slowing technological progress and economic growth. The following discussion examines each of these two concerns in greater detail and concludes that they do not provide a sound basis for either blocking or delaying statutory protection for databases. When compared to actual practices in the database industry, both concerns appear to be overblown or based on misunderstanding, and to the extent that these concerns are warranted, they can be addressed by appropriate regulations of statutory protection adopted by the Congress.

#### 4.1. Concerns about Market Power

**Economists have long recognized that the credible threat of a new market entrant is a powerful constraint on the ability of firms to exercise their potential market power.**

According to economic logic, an industry like the database industry with high upfront and low incremental costs of production has the preconditions for first-mover advantages and market power. This presumption is even stronger for so-called "niche" markets—like many specialized database markets—that may be large enough to justify the upfront costs of developing a specialized product but not large enough to support more than one or two competitors. Such analytical arguments, buttressed by anecdotal evidence, have led some skeptics of statutory protection for databases, such as the National Research Council and Professors Reichman and Samuelson, to conclude that "the market for commercially distributed databases is almost universally characterized by a distinct absence of competition" and that "the private database industry is largely characterized by niche marketers who supply and dominate specific market

segments."[30] These conclusions are not supported by either economic logic or empirical evidence.[31]

A large part of the database industry, including many of the most commercially significant databases, operates in an intensely competitive environment. Perhaps the largest single category of databases—measured in terms of volume of information and market size—involves financial data. There are four major players—Dow Jones, Reuters, Bloomberg, and Bridge (formerly Knight Ridder)—and thousands of smaller database providers, all serving the multi-billion-dollar demand for financial and commercial information. Interestingly, Bloomberg provides a good historical example of a firm that initially provided data only about a niche market—"sinking funds"—but has grown into a billion-dollar-a-year firm supplying a wide range of databases and news services.

Even the existence of a small number of firms serving a particular market does not mean that those firms exercise significant market power. First, as mentioned earlier, the database industry is characterized by differentiated products. Competing firms rarely supply the "same" database. Rather they compete on a range of fronts: selection of data; convenience; search engine; ease of use; and price.

Second, concerns about monopoly power in niche markets overlook the potential for substitution across databases that draw on somewhat different data sets. In many areas of research, for example, it is possible—and often wise—to investigate important issues and test general propositions with several alternative data sets. In addition, with the profusion of freely available information (for example, on the Internet) and powerful computers and computing tools, database makers face competition worldwide from competitors *and end-users* alike.

Third, and most important, in a market economy, firms prosper by supplying what their customers want. Firms that have satisfied customers get repeat business; those that do not tend to fail. Firms cannot "rest on their laurels" for fear that new competitors may come along and take away their business. Consequently, when one sees a niche database market supplied by a single firm, that may be evidence that the firm is doing a good job serving the needs of the market at reasonable prices.

Economists have long recognized that the credible threat of a new market entrant is a powerful constraint on the ability of firms to exercise their potential market power. The relevant public policy question, of course, is whether such a threat exists in a particular marketplace. Scholars who are concerned about the potential for market

---

[30] Recihman and Samuelson, 1997, p. 70.

[31] The anecdotal and survey evaluations of market power cited by the National Research Council are based on a count of how many firms compete to supply each database in specific database markets, without accounting for the different size of these markets in terms of revenues or sales. This procedure gives disproportionate weight to small database markets and is therefore biased toward showing a lower level of competition than that which characterizes the industry as a whole.

power abuse in the database industry emphasize the existence of high upfront costs and limited market size, especially for specialized products, as barriers to new market entrants. But over time, the shift toward electronic databases may well reduce some of the upfront costs of entry, as the prices of hardware, software, and communications technologies continue to fall. And at the same time, as these technologies spread broadly through educational, commercial, and family life, the demand for electronic information products is likely to expand, creating larger markets and luring additional entrants into database production. Moreover, because, as already noted, database producers must maintain in-house experts and technical staff and equipment to improve existing databases and keep pace with changing technological requirements, it may prove cost effective for them to dedicate these resources toward support of more than one database (especially if it is in an adjoining field for which their experts and staff already have expertise).

Finally, it may well be less costly for a new entrant to replicate an existing database than it was for its original producer to develop it, in which case the fixed costs of entry for the second firm are lower than they were for the first. There are three reasons why this is likely to be the case. First, most databases contain citations to the underlying sources from which they were developed. The original compiler had to track down these sources from among all of the possible sources where the data might reside. Much of this research may involve "blind alleys" and "false starts." A "second comer" can avoid these pitfalls and go directly to the sources identified in the existing database.

Second, the original developer of a database may have to spend a substantial amount of time and effort trying to track down data from a variety of potential sources, only to find that the information available from scattered sources is insufficient to yield a reliable data series. It may therefore cost the original developer a substantial amount to yield the "negative result" that certain desirable data are simply not available. A "second comer" can learn from the mistakes of the first. In the absence of the first firm's effort, the potential competitor would have to scour the record, determine what is and what is not available, and choose which data series to present. The incumbent firm has already done all that, and the new entrant is able to consider the results of those efforts.

Third, potential competitors can observe which databases are successful and which fail and choose to replicate only the former. In contrast, original developers of databases, like book publishers, movie studios, and record companies, often do not know how much demand there will be for a particular product until they develop and market it. Sometimes a database is more successful than anticipated, sometimes less—and often developers rely on the profits they make from their "hits" to cover the costs of their "flops." New entrants do not face the same degree of market uncertainty—they can take advantage of the "marketing experiment" performed by incumbents without themselves incurring the same risks.

Even if the first entrant does not significantly lower the costs of entry for new competitors in any one of these three ways, the possibility of new entrants exists as long as the underlying information contained in a particular database is available and can be replicated. That is why the proponents of statutory database protection have been careful to maintain that it should apply only to the database, not to the underlying data. It seems reasonable to argue that as long as potential new entrants—or potential users—can go to the same underlying data sources and replicate the information in a particular database, copying substantial parts of the original database should be precluded by statutory protection. How valuable the investment in the original database will be then depends on market demand, and if it proves to be sufficiently attractive, new entrants will be encouraged to enter the market and able to do so because they can turn to the original data.[32]

Moreover, as long as the underlying data can be replicated, potential buyers of database products can decide to buy the database from its original developer or collect the data themselves. Of course, one might object that this means that subsequent competitors or users will have to "duplicate" the work of the original developer and that such duplication of effort is wasteful. But this argument is mistaken for two reasons. First, it is likely that the second comer will innovate and improve on the first database in the course of replicating it. Second, concerns over unnecessary duplication of effort overlook the fact that second comers or users always have another choice—they can negotiate with the original developer for the right to use his product.[33]

If a mutually acceptable agreement is feasible, the parties will have an incentive to reach it so as to save the costs of replicating the data. For the second comer, the incentive is to save the costs of the duplicative effort. And the threat by the second comer to make that effort should the original developer fail to agree to terms will prevent him from charging an exorbitant price. Conversely, the ability of the original developer to compel the second comer to make good on that threat will in turn enable the original developer to extract a payment from the second comer that will help recoup part of the original investment necessary to develop the database in the first place. In short, the parties have a strong incentive to negotiate a more efficient solution if there are in fact mutually acceptable gains from trade.[34]

---

[32] In rare instances, data may not be easily replicable. We discuss these situations in section 4.2.

[33] We must be careful in saying that database protection is acceptable on the grounds that "competitors can always choose to replicate the work of the original database compiler if they want to compete." We protect a database against copying not because it can be replicated, but rather because denying protection diminishes incentive to develop the database in the first place.

[34] As an illustration, database producers may negotiate with potential competitors who are interested in licensing a database and incorporating it in a competitive product. The database producer will try to negotiate a price that reflects his assessment of the value of the resulting competition product in the marketplace and the likely decrease in revenue from the original product. In the absence of a successful contract negotiation, the potential competitor still has the opportunity to create his new product by replicating the data he needs from its original sources.

### 4.2. Competition and Data Replicability

**In most cases a potential entrant can get data from the same sources as the original firm, in which case there is no public policy need to allow the new entrant to free ride on the original firm's investment.**

As the preceding discussion indicates, the possibility of replicating the underlying data in a database is a key factor affecting the potential for market entry of new competitors. This observation naturally leads to the question: which kinds of data are replicable and which are not?

In fact, the underlying data in most databases is replicable. The clearest examples involve data collected by classic "sweat-of-the-brow" efforts from publicly available sources. A large number of databases are of this sort, including databases of historical information, many scientific databases, and databases that rely on government information such as census data, economic statistics, weather reports and readings, and indexes of government decisions and government actions such as court decisions. The fact that many commercially significant databases build on government information means that the government itself can encourage competitive database markets by making the data it generates publicly available on a non-exclusive basis. And in those cases in which the government decides to contract its data collection, maintenance, storage, verification, or dissemination activities to private firms, it should also require as part of the contract's conditions that the data be made publicly available on a non-exclusive basis to all users. In this case, the government's fee to its private contractor would cover costs plus whatever profit is negotiated as part of the contract, and the resulting data would then be made available to users in its raw form at incremental cost.[35] Of course, this leaves open the possibility that the contractor could develop his own database from the underlying data and charge a different and higher price for this product if the market values its additional services or features.

In fact, even when the government makes its data publicly available to a database developer and to end-users at the same price, individuals often choose to acquire their data from the private sector because of the value-added services such firms provide. Many businesses prefer to get government statistics from firms like DRI-McGraw-Hill rather than directly from the government because the private providers update and verify the data on a continuous basis and provide it in a readily usable electronic format. In California, for example, the state government has been putting much state-generated information on-line at no charge, competing with many private information

[35] The NRC has pointed to the US government's privatization of the provision of Landsat data as an example of the problems associated with granting statutory protection to databases. The NRC correctly points out that, following privatization, prices for Landsat images increased more than ten-fold, substantially impacting scientific users. But in our view these problems resulted from the manner in which the privatization contract was written and interpreted, not from the privatization effort itself.

suppliers that also provide government information to end-users and charge a fee for doing so. The increasing competition from the state government has caused private firms to introduce new value-added services to keep their customers satisfied. In the words of the leadership of one private firm, "We want to differentiate our product from the public-domain product, so people can make a choice of whether to use a free service or a paid service. The data is the data. We believe the difference is the accuracy, timeliness, ease of use and search, and other feature capabilities we can provide."[36]

What about statutory protection for databases that build on data that is not publicly available and hence cannot be easily replicated? Here several different situations should be distinguished. Sometimes a database relies on privately generated data that cannot be precisely replicated but for which comparable data would be available at comparable cost to competitors willing to make the necessary investment.

For example, during the 1980s supermarkets installed checkout devices that scanned bar-code labels on grocery items, and the A.C. Nielsen Company began to collect this data. Nielsen contracts with various supermarkets to supply it with scanner data. It then sells its ScanTrack reports about pricing, sales volume, and market share based on that data to grocery manufacturers. As we understand it, Nielsen typically pays the participating supermarkets a fee and provides them with copies of the reports in exchange for providing the data.

Our understanding is that these contracts between Nielsen and the selected supermarkets are exclusive: the stores agree to supply scanner data only to Nielsen. But other firms could and have entered into similar contracts with other stores to collect similar data. We know of at least one firm, IRI, which also provides scanner-based market data.

Of course, the data collected at other stores will be similar, but not identical. Nielsen might report that a particular detergent had a 17.4% market share in March in Nielsen-scanner-equipped stores; IRI might discover that the same detergent's market share in IRI-scanner-equipped stores during the same month was 18.7%.

That is, it may not be possible for Nielsen's competitors to *precisely* "replicate" Nielsen's data. But competitors can get "just as good" data from their own efforts. Consequently, there is no economic justification for allowing competitors to copy Nielsen's data merely on the grounds that it is not possible to "replicate" the data exactly.

Our next category involves privately generated data that cannot be replicated because it relies on proprietary information developed competitively. For example, suppose a pharmaceutical manufacturer develops a new compound, and runs scientific tests to

---

[36] Francis Bremson, Director of Marketing and Sales for Legi-Tech, quoted in Mitchell Benson, "State Web sites offer firms competition," Wall Street Journal, May 14, 1997, p. CA1.

determine its efficacy in treating a particular disease. Or suppose an electronics R&D house develops a new semiconductor manufacturing process and tests that process to determine product yields. These tests clearly generate databases of scientifically- and technologically-valuable information. And it would be difficult for competitors, without access to the compound or the new process, to replicate that data.

As a pragmatic matter, it is unlikely that these databases would be commercially sold; information of this sort is commonly closely-held as a trade secret. But if such a database were sold, there would be no economic rationale for allowing others to copy and sell it in competition with the original compiler, even though as a practical matter it would not be possible for the competitor to "replicate" the data contained in the database.

Another category involves historical data that cannot currently be replicated, but could have been at the time it was collected. Suppose that Smith compiled a city directory of Boston in 1985, by pounding the pavement. Today, in 1997, it is difficult if not impossible for Jones to try to replicate that database. People have moved, businesses have shut down, memories have faded; the information is no longer available. Does this justify Jones in copying the information in Smith's directory without Smith's permission (assuming that Smith's database is still protected)? We do not think so. (Of course, once the statutory protection on Smith's 1985 database expires, say in 2010 if database protection lasts for 25 years, Jones may reprint the database at no charge.)

This example shows that we must be careful in saying that database protection is acceptable on the grounds that "competitors can always choose to replicate the work of the original database compiler if they want to compete." In our view, the actual justification is a bit more complex. We protect Smith's database against Jones' copying, not because Jones can (currently) replicate it—in our city directory example, Jones cannot replicate in 1997 the data collected in 1985 by Smith—but rather because denying Smith protection diminishes Smith's incentive to develop the database in the first place.

If adequate protection is not provided for databases whose contents are not contemporaneously replicable, there will be little incentive to try to record the sorts of data that *cannot* be measured again—precisely the sorts of data that should be collected while it is still possible to do so.

Our last category involves privately generated data that others cannot replicate at comparable cost because of preferential governmentally established monopoly access to a different but related market. The main example we have in mind here is the telephone listings that were at issue in *Feist*.

Phone numbers are arbitrarily assigned identifying data, privately generated with no additional effort (the phone company assigns a number as a provision of service) in the course of operating a local-monopoly business. The government compels the phone

company to print phone directories so that this information, though privately generated, is publicly available.

According to our logic, the factual situations of the *Feist* case are in reality much closer to the kinds of concerns addressed in the antitrust law under the rubric of so-called "essential facilities" than they are to the kinds of concerns raised by a typical "database piracy" case. For the vast majority of databases, there is no credible essential facility claim. Either a potential entrant can get the data from the same sources as the original firm, in which case there is no public policy need to allow the new entrant to free ride on the original firm's investment, or the original firm generated the data itself in a non-monopoly context and others are equally free to generate their own data. When data is generated by a government-created monopolist, it is not appropriate to allow the monopolist to control database products building on that data. Such cases can be avoided by a policy that these sorts of data, collected by a government-created or government-sanctioned monopolist in order to provide its service, should be made available to other users.

**5. Concerns about Preferential Access**

> **The question of whether particular categories of users should get "preferential access" to information contained in privately produced databases is not the same question as whether the producers of these databases should be afforded adequate statutory protection.**

Many skeptics of statutory protection for databases are primarily motivated by their concern that it will restrict access of such preferred users as scientists and academics to essential data. Numerous exceptions and limitations to property rights in the copyright law encourage the use of protected property by such users for socially valued functions such as teaching, research and library activities. In addition, US copyright law contains a general "fair use" exception for purposes such as criticism, comment, news reporting, teaching, scholarship, and research. Neither the EC Directive nor HR 3531 contained any language indicating similar exceptions in statutory protection for databases,[37] and this caused concern, especially in the scientific community, that such protection could harm important activities. A recent report by the National Research Council[38] articulates these concerns. Although we are sympathetic to the needs of the scientific and educational communities—we either are or have been members of these communities for most of our adult lives—we believe that effective statutory protection for databases can be drafted to respect these needs. Indeed, we believe that such

[37] However, HR 3531 contained a number of provisions that had much the same effect, even if it did not invoke the words "fair use."

[38] Bits of Power: Issues in Global Access to Scientific Data, National Academy Press, 1997.

protection will be beneficial to these communities because it will provide the market incentives necessary to maintain healthy private investment in databases over time.

The question of whether particular categories of users should get "preferential access" to an economic good or service is not the same question as whether ownership rights over that product should be adequately protected. In addition, even if as a matter of public policy, certain kinds of users should receive preferential access, there remains the public policy issue of how that access is best achieved. A subsidy financing the purchase of a necessary good or service by preferred categories of users is the most direct form for realizing this objective. Allowing them to take the amount they want of a good or service from those who supply it without paying for it is only one particular form of subsidy—a subsidy in kind that is financed in essence by the suppliers and the other paying customers they serve. Usually such an in-kind subsidy proves considerably less efficient than a direct subsidy from the government. This conclusion can be illustrated by an example of preferred access drawn from a non-database setting.

The Women-Infants-Children (WIC) program is financed by the federal government and administered by the USDA. It provides mothers and their infant children with coupons that can be redeemed at participating retailers for milk, infant formula, and other specific staple food products like peanut butter and orange juice. The idea behind the WIC program is that early childhood malnutrition is clearly detrimental to both children and society and that one effective way to combat this problem is to subsidize the purchase of nutritional foods by those too poor to afford them. Several independent studies have concluded that the WIC program is both successful and cost-effective, and it enjoys bipartisan Congressional support.

But the fact that the WIC program has a sound public policy purpose does not mean that the best way to finance adequate nutrition for poor children is to allow their mothers to take milk and other food products from grocery store or other suppliers. Such an in-kind subsidy provided by these suppliers would tend to reduce their incentive to supply and increase the prices they charge to their paying, non-preferred customers. In addition, such an approach would tend to encourage waste or excessive use on the part of the subsidized population. In contrast, the WIC program provides direct subsidies that both regulate the level of usage by the preferred population and encourage producers to supply more, not less, of their output to the benefit of all users.

Similar insights apply when the "preferred class" of users consists of scientists, students, and academics rather than poor women and their children, and when the products in question are things like scientific equipment and databases. Consider the case of scanning electron microscopes. They are niche products with few suppliers, and virtually all of their customers are either academic or commercial scientists. Many scientists who would like to use electron microscopes for their research and educational work are unable to afford them, and government grants may provide financing—that is, monetary subsidies—to purchase such equipment. As far as we know, there have been no public policy proposals for an in-kind subsidy scheme

whereby such preferred users could simply take equipment from their suppliers at zero price. Nor would such proposals make good economic sense.

Some may argue in response that information is unlike milk or microscopes in that it is non-rival in use—that is, the use of information by some non-paying users will not reduce the supply of information available to paying users. But this argument overlooks the fact that in the absence of adequate property protection and rates of return, the supply of information produced by private companies will be reduced over time as investment levels fall. *In short, while information may be non-rival in use at a point of time when that information has already been developed, it is not non-rival in use over time—those who use existing information at zero price can adversely affect the growth of information over time to both their own detriment and to the detriment of paying users.*

We have no doubt that the question of access to data by the scientific and educational communities is a vitally important one for society as a whole. We also recognize that inadequate funding can deprive such preferred users of the data they need to do their work, thereby depriving society of the benefits of technological progress and educational attainment. Another related issue is access to information for such purposes as news reporting, commentary and criticism. But we do not believe that in-kind subsidies taken from database providers by such users in the form of unauthorized copying for unauthorized purposes are the appropriate remedy for addressing these valid concerns. Indeed, we believe quite the contrary—in order to provide conditions for a healthy and competitive database industry that will serve the needs of all users, adequate statutory protection is required. And as our earlier discussion indicates, the natural interest of database providers in broadening their audience promotes the practice of price differentiation whereby those who are able to pay more sometimes implicitly subsidize those who are not.

The need for adequate funding for science and education—to support the use of equipment, publications, software and data—is an important public policy challenge. Some of these funds come from government, some from private contributions, and some from the private sector in the form of company-sponsored support for research both within and outside of universities. But this challenge is logically distinct from the challenge of providing an appropriate environment for protecting the rights of private database producers.

This challenge is also logically distinct from the question of whether the government is putting adequate resources into the development and dissemination of those kinds of information of which it may well be the low-cost producer—information like statistics, weather data, space exploration data, and court opinions. Many of the concerns the scientific and educational communities have expressed in the debate about statutory database protection are in fact concerns about whether government spending on such information will be maintained at the necessary levels or whether the government will "privatize" these activities in ways that will increase their cost to data users.

## 6. Some Legal Concerns about Appropriate Statutory Protection for Databases

So far the discussion has focused on the economic logic for statutory protection for databases. We now turn to some questions about how best to provide such protection: How long should protection last? What mode should it take? What should be the scope of protection?

### 6.1. Duration of Protection

**The shorter the period of protection, the greater the incentive of producers to set high prices to try to recoup their investment during the allotted time.**

Under current US law, patents last 20 years from the date of filing; copyright protection lasts for the life of the author plus 50 years (or for 75 years from publication for works created by entities). The EC Database Directive protects databases for 15 years. HR 3531 proposed a 25-year term of protection.

From an economic perspective, it is often difficult to determine how long protection should last. Ideally, one might want the length of protection (and/or the scope of protection) to vary from database to database, but such a system would be totally impracticable. But setting a single term for database protection, applicable across all databases, presents a complicated task.

Finally, in thinking about the appropriate duration for database protection it is important to recognize that the shorter the period of protection, the greater the incentive of producers to try to set high prices to recoup their investment during the allotted time. Public policy makers should consider the likely relationship between duration of protection and firm pricing strategies in drafting statutory protection for databases.

### 6.2. Updating and protection

**Database providers spend hundreds of millions of dollars a year updating their existing databases, and the newly-updated databases also need protection.**

Up-to-date databases are clearly valuable—often extremely valuable. Stock traders need up-to-the-minute (even up-to-the-second) information about stock prices. Indeed, this allows providers of stock price information to "price differentiate." They charge a much higher price to those who need real-time stock price data than they charge to those who are willing to have the information delayed fifteen minutes. (Those willing to wait until the next day can buy it in the newspaper for a quarter.) But even in less time-sensitive fields, up-to-date information is almost invariably preferred to less recent and less complete information.

Consequently, database providers spend hundreds of millions of dollars a year updating their existing databases to add up-to-date information. The newly-updated databases also need protection.

Some commentators have expressed concern that various proposals for statutory protection for databases will enable database providers to obtain "perpetual" protection for their databases merely by updating the information contained in the database on a regular basis.

This argument is no more significant when applied to databases than it is when applied to updated copyrighted material (such as new editions of books or reference works). Many encyclopedias or other reference works are updated regularly. Each new edition is copyrighted, and the copyrights on old editions expire overtime. This does not provide "perpetual" protection. A good example is the CRC Press' Handbook of Chemistry and Physics, known to all science students, which has been published since 1914. Each new edition is copyrighted, and the copyrights on old editions expire over time.

We believe that a similar argument applies to proposals that would allow database providers to obtain protection for both an original database product and updated versions. Assume that databases are given 25 years of protection (as was proposed in HR 3531). The 1997 edition of the database would become available for copying by competitors in 2022. If in 2022 there were a market for the 1997 edition, a potential competitor could make a copy of it and compete with its original developer.

But what happens if an updated version of a database contains a substantial amount of information that was available in an earlier edition as well as a substantial amount of new content? Would the statutory protection accorded to the updated version extend to all of its contents—old and new alike—or merely to its "new content," as the copyright law does for printed matter? Since the point of statutory protection is to protect investment in the creation, verification, maintenance and dissemination of information, such protection when extended to updated products, in principle should apply only to new content and to the additional investment required to verify, maintain and disseminate old content included in new editions. Once the protection on a database runs its course, users should be allowed to copy its contents.

The problem with this approach, however, is that some or many of the old entries contained in an old database may have been re-verified or reorganized and possibly revised as the result of additional investment to produce the new edition. And there is likely to be no practicable way for the user to distinguish between old information in the new edition that has gone through such a process and old information that has simply been copied by the producer into the new version of his product. One way—and perhaps the simplest way—to handle this complication would be to apply statutory protection to all of the elements of a new version of a database. In other words, users would be prevented from copying "old information" as well as "new content" from a new version of a database, although they would be allowed to copy much of "the same" old information from a version whose protection had expired.

This approach would of course require that once protection for an old version of a database ended, users could still get copies of it. In the case of books and copyrights, the availability of old editions is fostered by the fact that copies of them are lodged with the Library of Congress and the Copyright Office. In the case of databases, statutory protection could require database producers to make archival copies of their databases on a regular basis (perhaps annually) and store them with the Library of Congress or some other agency. The archive copies would then be available after the statutory protection period had expired.

### 6.3. Use vs. Replication: On Sufficient Statistics and Scientific Research

**The claim that increased protection will impede the ability to use databases for scientific and academic research rests on a misunderstanding.**

The statutory proposals for database protection typically allow users to extract and copy small portions of a database, but seek to protect database developers against the copying of "all or substantial portions of" a database. Some have suggested that this would improperly limit the ability of legitimate users to make use of the database. We do not agree.

Users of a database often *make use* of "all or substantial portions of" of the information it contains in the course of large-scale data analyses, especially statistical analyses. It is typically—indeed, one would be tempted to say, nearly universally—impossible to determine what was in all or substantial portions of the entire database, starting only from the sorts of analyses and results that users in fact make of databases. To illustrate this point, we will indulge in a brief technical discourse.

To take a simple example, suppose that the database in question consists of the daily noontime temperature reading at City Hall in San Francisco for 90 consecutive days in the summer of 1988. Users want these 90 data points, not for their own sake but because they can be summarized, evaluated, compared with other areas, etc. For example, one might want to compare the summertime temperatures in San Francisco and Portland. For this comparison, one might want to summarize the 90 San Francisco data points into a single number that preserves some (but not all) of the data, but in a more compact form, such as the (arithmetic) average noontime temperature. For practical purposes, this single number may be sufficient for the comparison at issue. (Obviously, for other purposes it would not.)

But the important thing here is that we have "compressed" the data (from 90 points down to a single "summary" figure) in a way that cannot be reversed. One cannot go from the fact that "the average noontime temperature in San Francisco was 73 degrees" to a complete listing of 90 separate daily temperatures. To take a stark example, the average of the three numbers (3, 5, 7) is 5, but so is the average of (0, 5, 10) or of (2, 6, 7) or of (5, 5, 5). Being told the average does not enable one to recreate the underlying data.

The key point here is that scientists and other researchers typically can use "all or substantial portions of" the data in a database for their scholarly research, formulate and test hypotheses, do statistical analyses, write up their conclusions, and report their results in academic papers and journal articles.

At most, they may need to make and temporarily store (e.g., in RAM) an electronic "working copy" of the relevant data in a format useable by their statistics software package (or other analytic tool). Raw data—especially large amounts of raw data—may be the lifeblood of scientific research, but one key goal of science is to search the raw data for patterns and explanations. Raw data is a tool, not a goal in itself; it needs to be used with other tools (such as search engines and statistics packages) in order to yield useful results. *Absent the ability to manipulate and analyze the data, scientists would have little use for it in the first place.* Under copyright law, it is not clear whether the user of a database may make such temporary "working copies" without permission of the copyright holder. However, new legislation protecting databases can disregard the technicalities of copying and can address the economically significant element—the right to *use* the database and thereby benefit from the database producer's investment and labor. Thus, database protection legislation can make it clear that such temporary "working copies" are lawful so long as they are not used by unauthorized persons or in an unauthorized manner.

Consequently, in our view, one objection voiced by several commentators to increased database protection—namely, the claim that increased protection will impede the ability to use databases for scientific and academic research—rests on a misunderstanding of the difference between use and improper copying.

### 6.4. On Different Forms of Legal Protection For Databases

**Under current law, the main legal paradigm used to protect a database developer's interest in a database is a property-rule regime, as reflected in the common term "intellectual property" to refer to patent and copyright law. But there are also elements of a tort-based liability rule regime, notably in the law of trade secrets. One can use hybrid systems combining elements of tort and property.**

There are a number of different legal paradigms that might be used to protect a database developer's interest in a database. Under current law, the main paradigm is a property-rule regime, as reflected in the common term "intellectual property" to refer to patent and copyright law. But there are also elements of a tort-based liability rule regime, notably in the law of trade secrets. One can use hybrid systems combining elements of tort and property.

There are really two distinct but interrelated questions here. The first is: with respect to any particular proposed use by a particular type of user of a particular set of information in a particular database collected and/or generated by a particular database supplier, what are the respective rights of the parties? In their pioneering 1972 study of alternative forms of legal protection, Calabresi and Melamed referred to this as the question of how to define the *entitlements* of the various parties.[39] The answer may clearly vary from database to database, from one type of use to another, and from one type of user to another (*e.g.*, end-user *vs.* competitor).

Once these entitlements have been identified, the second question then turns to: how are those entitlements protected? With respect to each particular proposed use of a particular database, there are five general alternatives:

(1) The use may be permitted. Others–whether generally, or limited to certain types of individuals or firms–may be free to use the database in this particular fashion without the permission of the original database developer–indeed, often despite the implied or express opposition of the developer.

(2) The use may be subject to a compulsory license. Others (again, some or all) may be allowed to use the database in this fashion, but they must pay a pre-established fee to do so. This has two main alternatives:

   (a) the fee may be set by governmental regulation of some sort; or

[39] Guido Calabresi and A. Douglas Melamed, "Property Rules, Liability Rules, and Inalienability: One View From The Cathedral," Yale Law Journal, (1972), pp. 1089-1129.

(b) the database developer may be required to set and post a fee, which it may set (with or without some governmental oversight as to the size of the fee); but once set, the same fee must be charged on a non-discriminatory basis to all comers (or, at least, all comers within the appropriate category).

Presumably, in this context, there must be some sort of penalty attached to using the database use without paying the requisite fee. Otherwise, the infringer is in a "heads-I-win, tails-I-break-even" situation: since detecting infringement is not certain, the infringer may avoid paying the fee entirely if not caught, so that, absent a penalty, it would always be cheaper only to pay once caught.

(3) The use may be subject to damages under a liability rule: if others (again, some or all) use the database in an unauthorized fashion, they must pay damages to the database developer, with the level of damages determined on a case-by-case basis.

(4) The use may be protected by a property rule: others (again, some or all) may not use the database in this fashion without the (pre-negotiated) permission of the database developer, and can be enjoined from doing so. Under this approach, prospective users must negotiate with the database developer for the right to use the database in a particular fashion, and thus must meet the developer's price or forego the proposed use. Under this approach, the developer would be free to set a high price if it chose to do so.

(5) Finally, the use may be protected by criminal sanctions. Unauthorized use may result in fines or other penalties (including confiscation and destruction of infringing material, and possibly imprisonment).

These alternatives emphasize that different forms of protection can be used for different entitlements.

### 6.5. On The Choice Between Liability Rules And Other Forms Of Protection

**The fundamental policy choice between a property rights approach and liability rules turns on whether negotiations among the parties are possible. When negotiations are not possible, liability rules have been developed to recompense owners for past infringement of their rights through court-awarded damages. When negotiations are possible, a potential user of property can negotiate with its owner about the terms and conditions for its future use.**

As a matter of practical reality, one cannot undo the past. Earlier violations of entitlements cannot be undone; at most, the legal system can award damages for past infringement. And those damages have to be set by a court; one cannot rely

exclusively on after-the-fact assertions by the database developer, since it has an incentive to claim that it would not have sold at any price, in an effort to "prove" that damages for past infringement should be extremely high.

That is, looking backward at past infringement, the only realistic approach involves liability for damages.[40] But this does not mean that databases should therefore be protected by "liability rules."

The fundamental policy choice between property and liability rules turns on whether negotiations among the parties *before* infringement occurs are possible. In some contexts, like auto accidents, it is infeasible for people to negotiate beforehand. As Calabresi and Melamed put it,

> If we were to give victims a property entitlement not to be accidentally injured we would have to require all who engage in activities that might injure individuals to negotiate with them before an accident, and to buy the right to knock of an arm or a leg. Such pre-accident negotiations would be extremely expensive, often prohibitively so. To require them would thus preclude many activities that might, in fact, be worth having.[41]

Database protection is different from accidents. Those contemplating copying or using someone else's database almost invariably know who developed the original database, and are perfectly free to negotiate for the right to copy or use it.

Those arms'-length negotiations will establish a market price for those rights. We believe it is both unwise and inefficient to substitute court-established damages figures for market-established prices. It would establish the equivalent of a "private right of eminent domain," allowing others to take what they wish from a database without the owner's permission, subject only to paying court-awarded damages. The administrative burden imposed on the courts from trying to substitute court-awarded damages for privately-negotiated prices would be severe.

### 6.6. Property Rights in Databases are Not "Exclusive" Rights to the Data

**Giving a database developer a "property" right in a database does not "exclude" others from replicating the underlying data from original**

[40] As discussed earlier, a penalty should be attached to past infringement to avoid a "heads-I-win, tails-I-break-even" scenario on the part of the infringer.

[41] Calabresi and Melamed, 1972, p. 1109.

**sources and using or selling their own version of a database product based on that information.**

Some critics have express concern that the E.C. Directive and the 1996 Bill would give an "exclusive property right." We believe this concern to be misplaced, as it appears to rest on what we believe to be a misunderstanding of the "exclusivity" granted by property rights systems.

There is no question but that a property right gives the owner the ability to exclude others from using her property without her consent. But the extent to that this "excludes" others depends on their ability to replicate that property.

Giving a database developer a "property" right in a database does not "exclude" others from replicating the underlying data from original sources and using or selling their own version of a database product based on that information. That is, generally there are no barriers to entry into providing competing databases.

It is instructive to compare databases and patents on this point. A grant of a patent gives the patent holder the right to exclude others from making, using or selling the patented product or process. This right extends even to others who have never been exposed to the patent. Thus if Jones gets a patent, he can prevent Smith from using the patented technology, even if Smith developed the same technology independently and without knowledge of Jones' patent or Jones' research.

In this regard, patent protection is significantly unlike the proposals for database protection. If Smith has never been exposed to the contents of Jones' database, Smith is perfectly free to develop a similar if not identical database. Even if Smith *has* seen Jones' database, Smith is free to develop a competing database, so long as Smith does not directly copy from Jones' database.[42]

Hence giving Jones a "property right" in a database is significantly different from giving Jones a "property right" in a patent. While both are termed "property" rights, the ability to preclude others from independent development of the same product are significantly different.

### 6.7. "Value-Added" Products

**There is a well-established system by which those who seek to provide "value-added" products negotiate with the original authors and publishers for the right to do so. The fees to be paid to the original authors and publishers are set by negotiation.**

[42] Jones may deliberately "seed" false information in his database or look for inadvertent errors that reappear as ways to determine whether or not Smith copied from his database.

Some have criticized existing proposals for database protection because they assert that database developers would have an incentive to deny others the right to produce "value-added" products of various sorts. Again, we believe that this rests on a misapprehension of the role of private negotiations in facilitating the development of value-added products.

Economists differentiate between substitutes and complements. Many "value added products" are complements to the original database. The customer may find the database easier to use, or more valuable, because she also has the complementary product. For example, an index to a database, or a better search engine, or a manual or instruction book explaining how to use it more effectively, enhances the value of the database itself. In a literal sense, they "add value" to the original database. Complementary products increase demand for the original database, hence the database developer has incentive to encourage the development and marketing of complementary products. There is no reason to think that granting a database developer statutory protection for its databases will discourage the development of complementary "value added products." In any case, many such complementary products do not even require licensing negotiation because they do not incorporate protected parts of a database.

At the other extreme, some "value added products" are economic substitutes for the original database. This is especially likely to be the case if the substitute product contains a substantial amount of the contents of the original database. Consumers may or may not prefer the substitute product over the original database. Adequate statutory protection will allow the original developer to negotiate a license fee from a competitor for the use of the original database to develop a value-added product.

In the copyright sphere, there are many examples of "value-added" products. For example, the authors and publishers of a popular book may be approached by others seeking paperback publication rights, or foreign translation rights, or the rights to adapt the book into a movie. Popular children's films and cartoon shows generate proposals for all sorts of "tie-in" merchandise, from T-shirts to lunchboxes to toys. All of these can be considered "value-added" products.

There is a well-established system by which those who seek to provide such "value-added" products negotiate with the original authors and publishers for the right to do so. The fees to be paid to the original authors and publishers are set by negotiation.

Imagine, instead, that we used a liability rule under copyright law to determine how much the author or publisher of a book could charge for (say) foreign translation rights, or paperback publication rights, or the rights to develop the book into a movie or a TV series. Under such a system, translators or paperback publishers or TV and movie producers and studios would negotiate with authors and publishers, but if they were dissatisfied with the price they were able to negotiate, they could take the matter to a court to determine what a "reasonable" price should be for those rights. Given

the thousands of books published in paperback each year, and the hundreds of books that might be developed into movies or TV series, litigation to set a "reasonable" price for these rights could swamp the court system.

In essence, using a liability system to protect databases would, again, amount to granting to second comers a private "right of eminent domain," which would allow them to make use of databases developed by others at a court-determined rate. The law, rightly in our view, refuses to allow private parties to exercise a right of eminent domain, either with respect to real property or with respect to intangible property. We see no reason to adopt such a system for databases.

### 6.8 Compulsory Licensing

**A flat compulsory licensing fee would allow competitors to "skim the cream" by just copying the most successful databases. "Compulsory licensing" proposals are a particular form of price control, this time over the "price" of access to the contents of a database.**

In a similar vein, some critics of statutory database protection have urged that database providers be required to license their databases to "second comers" and those who wish to supply "value-added" products. For example, Professors Reichman and Samuelson have proposed that others should be free to use all or part of a database on "payment of reasonable compensation according to a menu of user options vetted by the industry with user and government inputs."[43]

As noted above, compulsory licensing is similar to liability rules, in that both substitute prices determined by the courts or by the government for prices determined by voluntary negotiations. The difference is that a compulsory license fee is determined *ex ante*, before the taking, and at a general level, while liability-based damages rules are determined *ex post*, after the taking has occurred, and on an individuated basis.

We believe that such a proposal is impracticable. Implementing any such system, and making sure that the "reasonable compensation" and the "menu of user options" keep up with technological and market changes, is likely to be a daunting task. We find it implausible in the extreme that "the industry" would be able to *agree* on what constitutes "reasonable compensation." Firms that develop their own databases obviously have a very different perception of what is "reasonable" than firms that merely clone databases developed by others. But both are part of the "industry" as it exists today.

[43] Reichman and Samuelson, 1997, p. 147.

Any system of compulsory licensing rates that allows new entrants to "pick and choose" elements they want to incorporate is likely to lead to the classic "cream skimming" problem. Some databases are clearly more economically valuable than others. How would one implement a "menu" that recognized this disparity? As we noted above, developing and marketing databases is a risky enterprise; database publishers rely on the profits they make from the "hits" to cover the cost of the "flops." A flat fee would allow competitors to "skim the cream" by just copying the most successful databases. This would undermine the database publishers' ability to develop all kinds of databases, including those specialized scientific and technical databases for which there is relatively low demand and which are of particular concern to Professors Reichman and Samuelson.

Economists have for years objected to the distortionary effects of price controls of all sorts. In our view, "compulsory licensing" proposals are at heart just a particular form of price control, this time over the "price" of access to the contents of a database developed by another. We do not believe that the government is well-placed to set such prices, for the same reasons that price controls are generally inefficient.

**7. Conclusions**

As the preceding discussion has demonstrated, there are strong economic reasons for providing adequate statutory protection for the database industry. Such protection would serve the interests not only of database producers, but also of database users, including users in the scientific, educational and library communities. The creation, storage, verification, maintenance, updating and dissemination of information serve valuable economic and public policy functions—and they are not "free" whether they are performed by the government or private companies. Indeed, such activities often involve substantial upfront costs and considerable risk, since it may be impossible to predict their actual value until the resulting information products are available for use. These costs and risks may be especially daunting for the development of highly specialized databases that are likely to have limited applications in the commercial arena, at least in the short run, and that therefore may have to rely initially on demand from a limited number of scientific and academic users with limited ability to pay.

Without effective statutory protection, private firms will be deterred from investing in database production. The resulting shortfall in the provision of information will have adverse effects on the pact of technological progress, on the economy's growth potential, and on the very research and educational communities whom critics of statutory protection wish to help.

PREPARED STATEMENT OF CONGRESSMAN JOHN D. DINGELL

Chairman Stearns and Chairman Smith, the debate over whether adequate legal protections exist to protect current databases and provide incentives to the creation of new and more valuable databases is an extremely contentious issue that has been debated in our respective Committees for several years. Databases are essential to ensuring the rapid search and retrieval of the enormous amounts of facts and other forms of information that are available, especially through the Internet. In fact, the Internet not only provides access to already created databases, but it has helped spur a remarkable growth in the number of databases, such as movie directories and loan comparison charts. Between 1990 and 2002, the number of database entries in the comprehensive Gale Directory of Databases has increased 147 percent. Moreover, the amount of information contained in such databases has increased 363 percent.

I note that this explosive growth in the number of databases has occurred despite the claims of the proponents of the draft legislation that "no meaningful legal protection of databases currently exists." I find it dubious that companies would invest vast amounts of financial resources in developing new databases if legal protections were nonexistent.

In fact, significant legal protections already exist for databases. For example, the original selection, coordination, and arrangement of facts in a database are protected by copyright law. Additionally, databases already receive protection under the Computer Fraud and Abuse Act, the Digital Millennium Copyright Act, and various state laws such as trespass to chattels, breach of contract, and misappropriation.

Notwithstanding the legal protections available, limited gaps in current law may exist. If such gaps are found, it is incumbent upon Congress to take a focused legislative approach as we attempted to do in the 106th Congress with H.R. 1858, the "Consumer and Investor Access to Information Act of 1999." This bill was narrowly crafted to provide limited protection to database producers against wholesale misappropriation of their work. Importantly, it would also have allowed the public to continue to have unfettered access to facts that are in the public domain.

Unfortunately, the draft Database and Collections of Information Misappropriation Act of 2003 takes the opposite approach. It would create broad new rights for database owners and dramatically alter our current information policy. Much like its predecessors, the draft bill has serious flaws and would stifle the development of a robust electronic commerce marketplace. It would create a quasi-property right in facts themselves, granting the compiler of information an unprecedented right to control value-added, downstream uses of the resulting collection. It would also establish an unprecedented subpoena process that would undoubtedly lead to abuse.

I must caution those who support broad new protections for databases. Electronic commerce has prospered in the United States in part because of our basic information policy - that facts, the building blocks of all information products, cannot be owned. Facts are part of the public domain. They do not owe their origin to an act of authorship. It is important that facts remain available for everyone to use and that Congress does not legislate in a way that would restrict the public's access to facts.

**Representative John Shimkus**
*Joint Committee Hearing of Subcommittee on Commerce, Trade, and Consumer Protection and Subcommittee on Courts, the Internet, and Intellectual Property*
**Hearing:**
*"Database and Collection of Information Misappropriation Act"*

September 23, 2003

OPENING STATEMENT

Good Afternoon. Chairman Stearns and Chairman Smith, Thank you for holding this hearing.

I will be upfront with the Committee, as I continue to follow this issue. I am not convinced that this is a pervasive problem. Those in favor of this legislation need to show the committee widespread evidence of database theft that cannot be remedied through existing legal, technological, or business methods.

The relative free flow of information is vital to the advancement of every aspect of our lives. Laws limiting the use of information need to be balanced against the cost and benefits of allowing people to build upon existing published ideas.

I find it very telling that a broad coalition of scientists, educators, research communities, libraries, Internet companies and databases, financial publishers and service providers, and large corporate users of information have voiced their fear of and opposition to this Discussion Draft.

I look forward to hearing from the distinguished panel. I yield back my time.

○