

A CONTRACTUALLY RECONSTRUCTED RESEARCH COMMONS FOR SCIENTIFIC DATA IN A HIGHLY PROTECTIONIST INTELLECTUAL PROPERTY ENVIRONMENT

J. H. REICHMAN* AND PAUL F. UHLIR**

TABLE OF CONTENTS

I.	INTRODUCTION	317
	A. Countervailing Trends Affecting the Production, Distribution, and Use of Scientific Data.....	319
	B. Formal and Informal Data Exchanges.....	322
	C. The Sharing Ethos Under Stress.....	323
	D. Scope and Structure of this Article	324
II.	RESEARCH OPPORTUNITIES IN A DIGITALLY NETWORKED ENVIRONMENT SUPPORTED BY A ROBUST PUBLIC DOMAIN.....	325
	A. Government-Generated Data: A Birthright of the Public Domain.....	326
	1. The Non-Proprietary Principle	326
	2. Limiting Factors.....	329

Copyright © 2003 by J.H. Reichman and Paul F. Uhlir

This article is also available at <http://www.law.duke.edu/journals/66LCPReichman>.

* J. H. Reichman is Bunyan S. Womble Professor of Law, Duke University.

** Paul F. Uhlir is Director of International Scientific and Technical Information Programs at the National Academies.

The opinions expressed in this article are the authors' and not necessarily those of the National Academies.

The authors gratefully acknowledge partial support for their work on this article from the Center for the Public Domain, under Grant No. OPVT-4676, and the John D. and Catherine T. MacArthur Foundation, under Grant No. 02-73708-GEN. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the supporting organizations.

Draft versions of this article were also presented at workshops sponsored by the National Research Council, Washington, D.C., Sept. 4-6, 2002, the Intellectual Property Research Institute of Australia ("IPRIA"), Melbourne, on Nov. 28, 2002, and the REGNET Social Sciences and Law Program, Australian National University, Canberra, Australia, Dec. 2, 2002. We are grateful to the participants for comments and suggestions. We also wish to thank the following individuals for their helpful comments and advice on previous drafts of this article: Peter Arzberger, Geoff Bowker, Andrew Christie, Peter Drahos, Peter Eckersley, Brett Frischmann, Janet Hope, Maureen Kelly, Steve Maurer, and William van Caenegem. Finally, we are grateful for the research support provided by Troy Petersen and Meredith Zinnani.

Portions of this article will also be appearing in National Research Council, *Proceedings of the Symposium on the Role of Scientific and Technical Data and Information in the Public Domain* (forthcoming 2003).

B.	Government-Funded Data: Between the Research Commons and Commercial Applications	331
1.	The Zone of Formally Regulated Access to Data.....	333
a.	Contractually Reinforcing the Sharing Ethos	334
b.	Legal Rules that Support the Sharing Ethos	336
c.	Countervailing Policies: The Commercialization of Government-Funded Research.....	341
2.	The Zone of Informal Data Exchanges	343
a.	Normative Conflicts	344
b.	A Different Legal Regime.....	348
C.	Private-Sector Data as a Component of the Research Commons....	351
1.	Dissemination and Use of Scientific Data in Private Sector R&D.....	351
2.	Database Publishers.....	354
D.	Potentially Enhanced Role of Public Domain Data in a Digitally Networked Research Environment	356
1.	Digitally Pooled Data Resources.....	356
2.	Electronic Transformation Tools	359
III.	SCIENTIFIC DATA AS A PRIVATE GOOD: PRESSURES ON THE PUBLIC DOMAIN AND THEIR IMPLICATIONS	361
A.	Shifting the Public-Private Boundaries.....	362
B.	Pressures on the Research Commons.....	366
1.	Commoditization of Data in Public Science.....	367
a.	Reducing the Scope of Government-Generated Data.....	368
b.	Commercial Exploitation of Academic Research	370
2.	Intellectual Property, E-Contracts, and Technological Fences...371	
a.	Expanding Copyright Protection of Factual Compilations: The Revolt Against <i>Feist</i>	374
b.	The DMCA: An Exclusive Right to Access Minimally Copyrightable Compilations of Data?	376
c.	One-Sided Electronic Licensing Contracts.....	379
i.	Restoring the Power of the “Two-Party” Deal.....	380
ii.	The Proposed Uniform Computerized Information Transactions Act (“UCITA”).....	381
d.	New Exclusive Property Rights in Non-Copyrightable Collections of Data.....	383
i.	The E.C. Database Directive in Brief.....	385
ii.	The Database Protection Controversy in the United States	388
C.	Implications for Science: Disintegration of the Research Commons?	396
1.	Restricting Access to and Use of Scientific Data	396
a.	In Government.....	396
b.	In Academia	398
i.	The Formal Zone	399

ii.	The Informal Zone.....	404
2.	Broader Implications for the Innovation System	410
a.	Underestimating the Potential Social Costs	410
b.	A Market-Breaking Approach.....	413
IV.	A CONTRACTUALLY RECONSTRUCTED RESEARCH COMMONS FOR SCIENCE AND INNOVATION	416
A.	The Challenge to Science: Formulating a Response to the Legal and Economic Pressures.....	417
B.	Proposals for the Government Sector	419
1.	Adjusting Domestic Policies and Practice	420
2.	Bridging the Gap with Foreign Law.....	423
C.	Proposals for the Academic Sector.....	425
1.	Formally Regulated Data Exchanges	426
a.	Basic Recommendations.....	427
b.	Ancillary Considerations	430
c.	Licensing Government-Funded Data to the Private Sector ..	432
i.	Key Questions	433
ii.	Value-Adding Uses and Management Costs.....	435
d.	Inter-University Licensing of Scientific Data.....	439
i.	Policy Considerations	440
ii.	Structuring Inter-University Data Exchanges	441
iii.	Resolving the Paradox of Conditional Deposits	444
iv.	Other Hard Problems	448
2.	Informal Data Exchanges.....	453
D.	Proposals for the Private Sector.....	457
V.	CONCLUSION.....	461

I

INTRODUCTION

Factual data are fundamental to the progress of science and to our preeminent system of innovation. Freedom of inquiry, the open availability of scientific data, and full disclosure of results through publication are the cornerstones of basic research, which both domestic law and the norms of public science have long upheld.¹

The rapid advances in digital technologies and networks over the past two decades have radically altered and improved the ways that data can be produced, disseminated, managed, and used in science and in all other spheres of

1. See NATIONAL RESEARCH COUNCIL, *BITS OF POWER: ISSUES IN GLOBAL ACCESS TO SCIENTIFIC DATA 2* (1997) [hereinafter *BITS OF POWER*]. Data are “facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors.” NATIONAL RESEARCH COUNCIL, *A QUESTION OF BALANCE: PRIVATE RIGHTS AND THE PUBLIC INTEREST IN SCIENTIFIC AND TECHNICAL DATABASES 15* (1999) [hereinafter *A QUESTION OF BALANCE*].

human endeavor.² As a result, these changes have given rise to a dramatic increase in the amount of data produced and have fostered unprecedented opportunities for accelerating research and creating wealth based on the exploitation of data.³ Every aspect of the natural world, from the sub-atomic to the cosmic, all human activities, and indeed every life form, can now be observed and captured through an electronic database.⁴ Whole areas of science are entirely data-driven, such as bioinformatics in molecular biology and the observational environmental sciences. All research increasingly depends on easy access to and use of data resources.⁵

Apart from the obvious technological advances that made these activities possible, much of the success of this revolution derives from the U.S. legal and policy regime that supports the open availability and unfettered use of scientific data.⁶ This regime, which remains among the most open in the world,⁷ has placed a premium on the broadest possible dissemination and use of scientific data produced by governmental or government-funded sources. This policy was traditionally implemented in several complementary ways: by expressly prohibiting intellectual property protection of all information produced by the federal government; by contractually reinforcing the sharing ethos of science⁸ through open data terms and conditions in federal research grants and contracts;⁹ by carving out a very large and robust public domain¹⁰ for non-copyrightable data;¹¹

2. BITS OF POWER, *supra* note 1, at 2; *see also* A QUESTION OF BALANCE, *supra* note 1, at 14-38.

3. *See generally* PAUL A. DAVID & DOMINIQUE FORAY, ECONOMIC FUNDAMENTALS OF THE KNOWLEDGE SOCIETY (Stanford Inst. for Econ. Pol'y Res., Discussion Paper No. 01-14, 2002), available at <http://siepr.stanford.edu/papers/pdf/01-14.html> (last visited Feb. 18, 2003).

4. J. H. Reichman & Paul F. Uhlir, *Database Protection at the Crossroads: Recent Developments and Their Impact on Science and Technology*, 14 BERKELEY TECH. L. J. 793, 812-13 (1999) [hereinafter Reichman & Uhlir, *Database Protection*].

5. BITS OF POWER, *supra* note 1, at 47-57.

6. *Id.* at 17.

7. *See* Peter N. Weiss & Peter Backlund, *International Information Policy in Conflict: Open and Unrestricted Access versus Government Commercialization*, in BORDERS IN CYBERSPACE 300, 307 (Brian Kahin & Charles Nesson eds., 1997).

8. For statutory waiver of copyright in government production, see 17 U.S.C. § 105 (2000). For the sharing ethos of science, see R.K. Merton, *The Normative Structure of Science*, in THE SOCIOLOGY OF SCIENCE 267-78 (R.K. Merton ed., 1973). *See also* PAUL A. DAVID, FROM KEEPING 'NATURE'S SECRETS' TO THE INSTITUTIONALIZATION OF 'OPEN SCIENCE,' 2 (Stanford Dep't of Econ., Working Paper No. 01006, 2001) [hereinafter DAVID, NATURE'S SECRETS], available at <http://www-econ.stanford.edu/faculty/workp/swp01006.pdf> (last visited Feb. 20, 2003); Michael Polanyi, *The Republic of Science: Its Political and Economic Theory*, 1 MINERVA 54, 59-79 (1962); BITS OF POWER, *supra* note 1, at 17-19, 21-22. For the environmental sciences perspective, see generally NATIONAL RESEARCH COUNCIL, ON THE FULL AND OPEN EXCHANGE OF SCIENTIFIC DATA (1995) [hereinafter FULL AND OPEN EXCHANGE] and NATIONAL RESEARCH COUNCIL, RESOLVING CONFLICTS ARISING FROM THE PRIVATIZATION OF ENVIRONMENTAL DATA 15-19 (2001) [hereinafter RESOLVING CONFLICTS], regarding scientists' views on the need for full and open access to environmental and earth science data.

9. *See infra* Part II.B.

10. We define "public domain" information as sources and types of data and information whose uses are not restricted by statutory intellectual property ("IP") laws and other legal regimes and that are accordingly available to the public for use without authorization. For analytical purposes, information in the public domain, including scientific data and information, may be divided into three major categories:

or by applying other immunities and exceptions that favor science and education to intellectual property rights that otherwise protect collections of information.¹²

A. Countervailing Trends Affecting the Production, Distribution, and Use of Scientific Data

A second and opposing trend, however, is characterized by the progressive privatization and commercialization of scientific data, and by the attendant pressures to hoard and trade them like other private commodities.¹³ This trend is reinforced by the creation of new legal rights and protectionist mechanisms that are largely extrinsic to the scientific enterprise,¹⁴ but increasingly adopted by it. These include greatly enhanced copyright protection of digital information;¹⁵ new ways to control access to and use of digital data by contractual

-
- (1) Information that is not subject to protection under exclusive IP rights.
 - (2) Information that qualifies as protectable subject matter under some IP regime, but that is contractually designated as unprotected (for example, is transferred or donated to a public archive or data center, or is made available directly to the public, with no rights reserved). Typically, such material consists of scientific data collections.
 - (3) Information that becomes available under statutorily created immunities and exceptions, which is also important in this context although it does not constitute public domain information per se.

“Open access” may be defined as proprietary information that is made openly and freely available on the Internet or through other media by the rights holder but that retains some or all of the exclusive property rights that are granted under statutory IP laws. Open access may be provided by all types of public and private sector sources. Of course, public domain information may be provided freely through open access as well. By no means is all public domain information freely available, however, even though once accessed, it may be used without restriction. This article focuses primarily on scientific and technical (“S&T”) data in the public domain available through open access. *See generally* NATIONAL RESEARCH COUNCIL, PROCEEDINGS OF THE SYMPOSIUM ON THE ROLE OF SCIENTIFIC AND TECHNICAL DATA AND INFORMATION IN THE PUBLIC DOMAIN (forthcoming 2003) [hereinafter NRC SYMPOSIUM].

11. 17 U.S.C. §§ 102(a)-(b), 103(b) (2000); *Feist Publ'ns, Inc. v. Rural Tel. Servs. Co.*, 499 U.S. 340 (1991).

12. *See, e.g.*, 17 U.S.C. § 107 (2000) (fair use); § 108 (reproductions by libraries and archives); § 109(a) (first-sale doctrine); § 110(1) (face-to-face teaching activities); § 110(2) (educational broadcasts).

13. *See, e.g.*, Rebecca S. Eisenberg, *Bargaining Over the Transfer of Proprietary Research Tools: Is this Market Failing or Emerging?*, in EXPANDING THE BOUNDARIES OF INTELLECTUAL PROPERTY: INNOVATION POLICY FOR THE KNOWLEDGE SOCIETY, at 223-49 (Rochelle Dreyfuss et al. eds., 2001) [hereinafter Eisenberg, *Bargaining*] (stressing delays and high transaction costs impeding transfers of university-generated biotech research tools); Walter W. Powell, *Networks of Learning in Biotechnology: Opportunities and Constraints Associated with Relational Contracting in a Knowledge-Intensive Field*, in *id.* at 251, 263-65 (stressing “sea change in the focus of basic research” in life sciences owing to commercialization by universities of basic science discoveries, increasingly under exclusive property relationships).

14. *See generally* COMMITTEE ON INTELLECTUAL PROPERTY RIGHTS AND EMERGING INFORMATION INFRASTRUCTURE, NATIONAL RESEARCH COUNCIL, THE DIGITAL DILEMMA: INTELLECTUAL PROPERTY IN THE INFORMATION AGE 96-122, 152-98 (2000) [hereinafter DIGITAL DILEMMA].

15. Digital Millennium Copyright Act of 1998 (DMCA), 17 U.S.C. §§ 1201-1203 (2000). Irrespective of the DMCA, federal appellate courts have begun to broaden copyright protection of low authorship publications. *See infra* text accompanying notes 286-87.

restrictions that are technologically enforced;¹⁶ and the enactment of proposals for novel intellectual property rights¹⁷ to protect collections of data.¹⁸

These new legal rights and mechanisms are being promoted by certain information industry conglomerates because of economic opportunities for the private exploitation of new digital information resources and as a legal reaction to a possible loss of control over certain proprietary information products in the digital environment.¹⁹ At the same time, the new laws pose the danger of disrupting the normative customs at the foundation of public science, especially the traditional cooperative and sharing ethos, by producing both the pressures and the means to enclose the scientific commons and to greatly reduce the scope of data in the public domain.²⁰ Viewed dispassionately, the need to reconcile these trends in a socially productive framework has become imperative, and the goal of such a reconciliation seems clear. A positive outcome would maximize the dissemination of scientific data in a quasi-public space where access and use for research purposes was ensured, without disrupting new opportunities for commercial exploitation of scientific databases in the private sector.²¹

Perhaps a single example at the outset will help to illuminate some of these developments. Under traditional assumptions, scientific researchers would

16. See, e.g., *ProCD, Inc. v. Zeidenberg*, 86 F.3d 1447 (7th Cir. 1996). See generally J. H. Reichman & Jonathan A. Franklin, *Privately Legislated Intellectual Property Rights: Reconciling Freedom of Contract with Public Good Uses of Information*, 147 U. PA. L. REV. 875 (1999); Symposium, *Intellectual Property and Contract Law for the Information Age: The Impact of Article 2B of the Uniform Commercial Code on the Future of Information and Commerce* (pts. 1 & 2) 87 CAL. L. REV. 1 (1999), 13 BERKELEY TECH. L.J. 809 (1998) [hereinafter *Intellectual Property & Contract Law*].

17. The U.S. House of Representatives' Committee on the Judiciary had introduced a series of bills modeled after Directive 96/9 of the European Parliament and the Council of 11 March 1996 on the legal protection of databases, 1996 O.J. (L 77) 2 [hereinafter E.C. Database Directive]. The most recent officially introduced version was H.R. 354, the Collections of Information Antipiracy Act (2000). See Reichman & Uhler, *Database Protection*, *supra* note 4, at 824. In 1999, the House Committee on Commerce (now called the House Committee on Energy and Commerce) introduced a more narrowly drawn version of database protection legislation based on unfair competition law principles: the Consumer and Investor Access to Information Act of 1999, H.R. 1858.

18. For detailed discussion of the E.C. Database Directive as it impacts science, see Reichman & Uhler, *Database Protection*, *supra* note 4. See also J. H. Reichman & Pamela Samuelson, *Intellectual Property Rights in Data?*, 50 VAND. L. REV. 51, 114-23 (1997). See generally Jane C. Ginsburg, *U.S. Initiatives to Protect Works of Low Authorship, in EXPANDING THE BOUNDARIES OF IP*, *supra* note 13, at 55, 68-72 [hereinafter Ginsburg, *U.S. Initiatives*]; J. H. Reichman, *Database Protection in a Global Economy*, 2002 REVUE INTERNATIONALE DE DROIT ECONOMIQUE 455-504 (2002) [hereinafter Reichman, *Database Protection in a Global Economy*].

19. See generally DIGITAL DILEMMA, *supra* note 14, at 51-58, 61-67.

20. See generally PAUL A. DAVID, THE DIGITAL TECHNOLOGY BOOMERANG: NEW INTELLECTUAL PROPERTY RIGHTS THREATEN GLOBAL 'OPEN SCIENCE' (Stanford Dep't of Econ., Working Paper No. 00-006, 2000) [hereinafter DAVID, DIGITAL BOOMERANG], available at: <http://www-econ.stanford.edu/faculty/workp/swp00016.pdf>; PAUL A. DAVID, WILL BUILDING "GOOD FENCES" REALLY MAKE "GOOD NEIGHBORS" IN SCIENCE? DIGITAL TECHNOLOGIES, COLLABORATIVE RESEARCH ON THE INTERNET AND THE EC'S PUSH FOR PROTECTION OF INTELLECTUAL PROPERTY, (Stanford Inst. for Econ. Pol'y Res., Discussion Paper No. 00-33, 2000) [hereinafter DAVID, GOOD FENCES], available at <http://siepr.stanford.edu/papers/pdf/00-33.pdf> (last visited Feb. 20, 2003).

21. For our proposals to achieve such a positive outcome, see Part IV of this article.

publish their findings together with the supporting data. Both the findings and the data would enter the public domain and become part of the scientific knowledge base. Traditional copyright principles have supported this result, as have the normative practices of the scientific community.²² Today, however, the fact that the data are published with the article, whether in print or digital form, may tell us little or nothing about the terms of their accessibility and their use by other scientists for follow-on work, for a variety of reasons.

First, as a growing commercial or cultural phenomenon, the data may have been conditionally deposited or imperfectly revealed at the time of publication.²³ Second, recent changes to copyright law make it possible to control online access to the supporting data, even though the data as such are technically ineligible for copyright protection.²⁴ Third, European states have adopted a new *sui generis* database right, which allows scientists to directly control access to and reuse of aggregations of facts, whether these have been disclosed as part of their research publications or made available as a separate database.²⁵ Bills to enact similar legislation have been introduced in the United States.²⁶

Finally, even disregarding these major changes in the underlying intellectual property regime, a combination of digital rights management technologies and standard-form contracts may enable publishers to impose limits on the redistribution and use of supporting data even after formal publication of a scientific article. This power of the “two-party” deal, which modern telecommunications technology has restored, would grow even stronger in the United States if either a strong database protection bill were enacted or a pending proposal for new and uniform contract laws regulating information products were more widely adopted at the state level.²⁷ If both types of laws were adopted, either the scientist or the publisher could retain maximum power to control subsequent access to and use of any compilation of scientific data (but not individual facts) otherwise disclosed to the public and previously in the public domain.

The foregoing list illustrates the very far-reaching revisions in the legal rules that have supported traditional modes of accessing and exchanging scientific data and databases. The resulting problems are further complicated by the

22. See *infra* text accompanying notes 78-113; BITS OF POWER, *supra* note 1, at 2 (regarding the normative practices of the scientific community); Reichman & Uhler, *Database Protection*, *supra* note 4, at 800-02 (discussing user-friendly rules of copyright law).

23. See, e.g., Ari Patrinos & Dan Drell, *The Times They Are A-Changin'*, 417 NATURE 589, 589-90 (June 6, 2002).

24. See Digital Millennium Copyright Act of 1998 (DMCA), 17 U.S.C. §§ 1201-1203 (2000); Ginsburg, *U.S. Initiatives*, *supra* note 18, at 61-67.

25. E.C. Database Directive, *supra* note 17.

26. See H.R. 354, 106th Cong. (1999) and H.R. 1858, 106th Cong. (1999). Previous versions of the Collections of Information Antipiracy Act introduced by the House Committee on the Judiciary included H.R. 2281, 105th Cong. (1998); H.R. 2652, 105th Cong. (1997), and The Database Investment and Intellectual Property Antipiracy Act of 1996, H.R. 3531, 104th Cong. (1996).

27. Reichman & Franklin, *supra* note 16, at 897-99.

changing ways in which the scientific community itself uses data and exchanges them on both a formal and informal basis.²⁸

B. Formal and Informal Data Exchanges

Traditionally, in “big science” projects, in disciplines such as physics, space, and earth sciences,²⁹ government science agencies play the predominant, controlling role in the collection and dissemination of data from large facility instruments. The agencies themselves collect the raw data, or they fund the production of data by academia through large, highly structured, and long-term research programs. Such data are then typically managed in well-organized data centers, where they are deposited on terms of open public domain access for the worldwide scientific community. Other science agencies also sponsor and fund research that makes use of these collected inputs. The public research that emerges is disclosed through peer-reviewed publications or applied to commercial endeavors. These publications or applications will then attract exclusive intellectual property rights—copyrights, patents, or, increasingly, *sui generis* rights.

What is noteworthy about this picture is the high degree of public control over the data that flow through the system. Because the government itself collects, funds, or disseminates so much of this data, it has a great deal to say about the rules of access and use that apply. It has thus promoted open access to research data as a public good, and through its use of public grants and contracts has reinforced the sharing ethos to which the scientific community traditionally subscribed. This ethos, in turn, fits comfortably within the underlying legal infrastructure, which typically distinguished between information as inputs, not subject to intellectual property rights, and aggregates of information bundled as outputs, which do attract such rights.³⁰

Of course, it would be wrong to paint too rosy a picture here or to assume that data in “big science” programs have flowed through and supported the public research system optimally and that the open access and sharing norms have been rigorously and uniformly enforced. Moreover, the areas of “small science” have never really fit within this sociological description of the data access system sketched above.³¹ Small science is being performed by individual

28. See discussion *infra* accompanying notes 65-157 and 183-210.

29. See generally BIG SCIENCE: THE GROWTH OF LARGE-SCALE RESEARCH (Peter L. Galison & Bruce Hevly eds., 1994) [hereinafter BIG SCIENCE].

30. See generally Reichman & Franklin, *supra* note 16, at 884-88 (“The Dual Function of Information in the Networked Environment”) (citing authorities).

31. See generally Stephen Hilgartner, *Access to Data and Intellectual Property: Scientific Exchange in Genome Research*, in INTELLECTUAL PROPERTY RIGHTS AND THE DISSEMINATION OF RESEARCH TOOLS IN MOLECULAR BIOLOGY: SUMMARY OF A WORKSHOP HELD AT THE NATIONAL ACADEMY OF SCIENCE, Feb. 15-16, 1996, 28-39 (1997) [hereinafter Hilgartner, *Access to Data*]; Stephen Hilgartner & Sherry I. Brandt-Rauf, *Controlling Data and Resources: Access Strategies in Molecular Genetics*, in INFORMATION TECHNOLOGY AND THE PRODUCTIVITY PARADOX (P.A. David & W.E. Steinmueller eds., 1998) [hereinafter Hilgartner & Brandt-Rauf, *Controlling Data*]; Sherry I. Brandt-Rauf, *The Role,*

investigators or small and autonomous research groups operating outside large, organized research programs, often with non-federal sources of funding. In these areas of science, data are generated in relatively small amounts, and single laboratories or investigators work independently. Nevertheless, the latter also find themselves dependent on the results of those other individuals in their field of inquiry, to which they have no guaranteed access. Here, the data exist in various twilight states of accessibility, depending on the extent to which they are published, discussed in papers but not revealed, or just known about because of reputation or ongoing work but kept under absolute or relative secrecy.

There are few government-controlled, public domain data centers in this type of research. The data are thus disaggregated components of an incipient network that is only as effective as the individual transactions that put it together. Openness and sharing are not ignored, but they are not necessarily dominant, either. These values must compete with strategic considerations of self-interest, secrecy, and the logic of mutually beneficial exchange.

In small science, what occurs is a delicate process of negotiation, in which data are traded as the result of informal compromises between private and public interests that are worked out on an ad hoc and continual basis. Small science thus depends on the flow of disaggregated data through many different hands, all of which collectively construct a fragile chain of semi-contractual relations in which secrecy and disclosure are pitted against a common need for access and use of these resources. In this sense, big science projects are more likely to be subject to formal data access regulations while small science research is more emblematic of informal data exchange practices.

C. The Sharing Ethos Under Stress

The picture painted herein of “big” and “small” science is overstated to clarify the basic concepts. The big science system is not free of individual strategic data transactions that are typical of the small sciences, and the latter often benefit from access to public repositories where data are freely and openly available, rather than through the ad hoc transactional process. However, the “brokered networks” typical of small science are endemic to all sciences, and access to data is everywhere becoming more dependent on negotiated transactions between private stakeholders. These outcomes are increasingly achieved at the expense of the public interest, and they may bypass the contributions to public data repositories and the sharing norms of the past.³² Moreover, in recent years the pressures to commoditize and privatize research results have extended farther upstream in the process, which affects the government’s attitude toward the creation and dissemination of its own data in support of research and other public interests.

Value, and Limits of S&T Data and Information in the Public Domain for Biomedical Research, in NRC SYMPOSIUM, supra note 10.

32. Cf. Powell, *supra* note 13, at 263-64 (stressing risks of undermining public science).

The pressures on access to scientific data from privatization and commercialization could, in short, be a big story even if the underlying legal infrastructure were to hold constant.³³ In reality, the massive changes in the legal regime already alluded to could have profound effects on access to and use of data in every scientific discipline, whether a part of “big” or “small” science, or even “basic” or “applied.”

The biggest threat to the integrity of the system is the extension of exclusive intellectual property rights to collections of data themselves, which poses a grave threat to the continuity of all those data exchange processes that scientists in the United States have long taken for granted. Moreover, this is not a battle that science can win, in the sense that there are no legislative fixes likely to stave off the many different legislative threats to the open exchange of data in the public domain. The pressures to commoditize are in many cases becoming too great, and the legal and technical protection instruments are now too varied and refined to be thwarted by ad hoc legal and technical responses. Any legislative remedies that might provide narrow exceptions for public research and education from an increasingly high-protectionist regime cannot resolve the major obstacles to the open availability and exchange of scientific data heretofore in the public domain. Rather, precisely because these pressures on the system are becoming so great, we contend that the scientific community can and must assert greater control over the management of its own data supplies.

D. Scope and Structure of this Article

The aim of this article is to promote discussion and a greater understanding within both the scientific and legal communities of the important trends and issues noted above. Part II examines the role and value of public domain data in scientific research and maps the way it functions today. It describes the legal and policy framework for government-generated and government-funded scientific data activities, the related sociological and normative structure of the scientific enterprise in the formal and informal contexts, and some of the opportunities for the exploitation of data in public research that are inherent in digital technologies and networks.

Part III begins by looking at the shifting public–private boundaries. It then documents many of the economic and legal pressures on scientific data traditionally in the public domain and the threats these changes pose to the established norms and practices supporting open access to and use of data in research.

Finally, Part IV demonstrates how the scientific community can manage its way out of these dilemmas, but only if it is willing to come to terms with real-world commercial pressures that will require some significant compromises to preserve an acceptable balance of public and private interests. Toward this end,

33. Cf. Arti Kaur Rai, *Regulating Scientific Research: Intellectual Property Rights and the Norms of Science*, 94 NW. U. L. REV. 77, 92-94, 109-15 (1999).

we propose a dual strategy, one that contractually reinforces the public domain for data that exists within the ambit of the federal government and another that contractually reconstructs a research commons for data (and other forms of information) in academia and the private sector. We argue that excessively rigid efforts to keep scientific data free of private control will end by yielding less and less data to the public domain, whereas a contractually reconstructed commons for data, while less pure in theory, will in practice make more data more accessible for research purposes in the long run. To make this strategy work, the funding agencies, universities, and scientific organizations must agree to a basic set of ground rules, with the goal of preserving the data commons for research purposes without impeding institutional actors or single researchers from enjoying the benefits of appropriate commercialization in the private sector.

II

RESEARCH OPPORTUNITIES IN A DIGITALLY NETWORKED ENVIRONMENT SUPPORTED BY A ROBUST PUBLIC DOMAIN

The producers of scientific data can be divided into three sectors: government agencies (primarily but not exclusively federal), academic and other not-for-profit research institutions, and private sector commercial enterprises. In the past, the research activities of the first two sectors—the government agencies and the non-profits—operated largely unimpeded by the classical intellectual property system. In fact, that system indirectly supported these activities by facilitating access to a robust public domain. Regarding operations in the private sector, where the objective is to commercialize data, few and relatively weak intellectual property rights have nonetheless supported private investment in a vigorous U.S. database industry that dominates the world market.³⁴ At the same time, this intellectual property environment has facilitated the informal exchanges of individual scientific researchers and has not unduly impeded research at either the government or university level, except insofar as licensing contracts supported by new technological protection measures have begun to cut back on pre-existing freedom of access and use.

If the pre-existing legal regime were to remain supportive, the role and value of public domain data could potentially be magnified many times over by

34. In 1998, sixty-three percent of databases were reportedly produced in the United States. Although the domestic database industry continued to expand, its share of the global output declined from a ratio of U.S. to non-U.S. databases of about two-to-one in the 1985-1993 period, to a ratio of three-to-two in 1998. Martha E. Williams, *State of Databases Today: 1999*, in GALE DIRECTORY OF DATABASES (L. Kumar ed., 1998). These statistics, however, do not include large numbers of government and academic databases that are not officially registered. See A QUESTION OF BALANCE, *supra* note 1, at 28; see also Cynthia M. Bott, *Protection of Information Products: Balancing Commercial Reality and the Public Domain*, 67 U. CIN. L. REV. 237 (1998); Stephen M. Maurer, *Across Two Worlds: Database Protection in the U.S. and Europe*, paper prepared for Industry Canada's Conference on Intellectual Property and Innovation in the Knowledge-Based Economy 8-21 (May 23-24, 2001) [hereinafter Maurer, *Across Two Worlds*] (comparing the U.S. database industry with that of other countries).

the advent of digital technologies and new research tools. In reality, however, this potentially enhanced role of public domain data in science is threatened by a confluence of economic, legal, and technological pressures described in detail in Part III.

A. Government-Generated Data: A Birthright of the Public Domain

The role of government in supporting scientific progress in general,³⁵ and its influence on the creation and maintenance of the research commons in particular, cannot be overstated. The U.S. government produces the largest body of public domain data and information used in scientific research and education. For example, the federal government alone now spends more than forty-five billion dollars per year on its research programs,³⁶ with a significant percentage of that money invested in the production of primary data sources, in higher-level processed data products, statistics, and models, and in scientific and technical information (“STI”), such as government reports, technical papers, research articles, memoranda, and other such analytical material.

1. The Non-Proprietary Principle

The United States—unlike most other countries—overrides the canons of intellectual property law that could otherwise endow it with exclusive rights in government-generated collections of data or other information. To this end, the Copyright Act of 1976 prohibits the federal government from claiming protection of its publications.³⁷ The bulk of the data and information produced directly by the government automatically enter the public domain year after year, with no proprietary restrictions.

The federal government also provides the greatest amount of funding for the production of scientific data by the non-governmental research community, as a significant part of that forty-five billion dollar investment in public research,³⁸ and many of the government-funded data collections it yields also become available to the scientific community through the research commons. Government-funded data follow a different trajectory from that of data produced by the government itself, however, and discussion of that topic is deferred to the next Part.

A number of well-established reasons support the policies that promote open access to and use of government-generated data, often at no cost to the public. The government needs no legal incentive to create the information; the taxpayer has already paid once for the production of a database or report and should not pay twice; transparency of governance and democratic values would

35. For an excellent overview of the role of the U.S. government in the domestic research system, see DONALD E. STOKES, *PASTEUR'S QUADRANT: BASIC SCIENCE AND TECHNOLOGICAL INNOVATION* (1997).

36. American Association for the Advancement of Science R&D Funding Update of Mar. 14, 2002, Table 2, at <http://www.aaas.org/spp/rd/prev03pt.htm> [hereinafter AAAS R&D Funding Update].

37. 17 U.S.C. § 105 (2000).

38. AAAS R&D Funding Update, *supra* note 36.

be undermined by limiting broad dissemination and use of public data and information; citizens' First Amendment rights might be compromised; and the nation generally benefits from broad, unfettered access to and use of government databases and other public information by all citizens to promote economic, educational, and cultural values.³⁹ It is primarily the latter justification, which involves many positive externalities and network effects from the Internet on the conduct of research and on our national system of innovation that is this article's focus of discussion.⁴⁰

The federal government's specific policies relating to scientific data activities date back to the advent of the era of "big science" following World War II, which established a framework for the planning and management of large-scale basic and applied research programs.⁴¹ Most of this research was initially conducted in the physical sciences and engineering, fueled largely by the Cold War and related national defense concerns. Although a substantial portion of this research was classified, at least initially, the default rule was that the research data and information produced by the government entered the public domain. This research model yielded a succession of spectacular scientific and technological breakthroughs and well-documented socio-economic benefits.

The hallmark of big science, more recently referred to as "megascience," has been the use of large research facilities or research centers and of "facility-class" instruments, which are most usefully characterized as observational or experimental.⁴² In the observational sciences, some of the most significant advances initially occurred in the space and earth sciences as offshoots of classified military and intelligence space technologies and of NASA's Apollo program. Notable examples of large observational facilities have included space

39. See Weiss & Backlund, *supra* note 7, at 300-05; A QUESTION OF BALANCE, *supra* note 1, at 52-58. See generally Henry H. Perritt, Jr., *Sources of Rights to Access Public Information*, 4 WM. & MARY BILL RTS. J. 179 (Summer 1995);

40. The policies and practices of state and local governments with regard to legal protection of their databases and other information are not as straightforward as those in the federal context. Section 105 of the 1976 Copyright Act does not expressly ban copyright claims in the works of non-federal government entities. 17 U.S.C. § 105 (2000). Some states have nonetheless enacted open records laws that prohibit protection of their government information, that encourage open dissemination to the public, and contain provisions analogous to the Freedom of Information Act ("FOIA"), 5 U.S.C. § 552 (2000). See, e.g., California Public Records Act, CAL. GOV'T CODE § 6253 (Deering 2003); 5 ILL. COMP. STAT. 140/3 (2002). There is no uniformity among the states in these areas, however, and there are many exceptions that allow state and local jurisdictions to protect some types of information generated by selected agencies, even in those states that have enacted open records laws. Consequently, some state and local agencies currently protect their databases and other productions under copyright and contract laws, and these agencies would likely make use of any additional intellectual property protection that new federal or state laws might provide. Nevertheless, most of the same policy reasons that support the public domain status of federal government information apply equally well at the lower levels of government and thus should exempt information produced by state and local governments from such protection. For an overview of state practice and policy implications, see Perritt, *supra* note 39.

41. See generally BIG SCIENCE, *supra* note 29.

42. See BITS OF POWER, *supra* note 1, at 58-61; see also OECD, EVALUATION OF THE OECD MEGASCIENCE FORUM; REPORT OF THE EXPERT PANEL (1998), available at <http://www.oecd.org/pdf/M000014000/M00014730.pdf> (last visited Feb. 13, 2003).

science satellites for robotic solar-system exploration, ground-based astronomical telescopes, earth observation satellites, networks of terrestrial sensors for continuous global environmental observations and global change studies, and more recently, automated genome-decoding machines.⁴³ Major examples in the experimental sciences have included facilities for neutron beam and synchrotron radiation sources, large lasers, supercolliders for high-energy particle physics, high-field magnet laboratories, and nuclear fusion experiments.⁴⁴

The data from many of these government research projects, especially in the last two decades, have been openly shared and archived in public repositories. Hundreds of specialized data centers have been established by the federal science agencies or at universities under government contract.⁴⁵

Scientific data and other kinds of information generated by the governments of other nations may also end up in the public domain and become available internationally.⁴⁶ Generally, however, the quantities are considerably smaller than the information resources generated by the U.S. government, and the

43. BITS OF POWER, *supra* note 1, at 58-61.

44. *Id.*

45. A few well-known examples of the government's public domain data archiving and dissemination activities include the NASA Space Science Data Center, <http://nssdc.gsfc.nasa.gov> (last visited Feb. 18, 2003); the National Oceanic and Atmospheric Administration's ("NOAA") National Data Centers, <http://www.nesdis.noaa.gov> (last visited Feb. 20, 2003); the U.S. Geological Survey's Earth Resources Observation Systems ("EROS") Data Center; <http://edc.usgs.gov> (last visited Feb. 18, 2003); and the National Center for Biotechnology Information at the National Institutes of Health, <http://www.ncbi.nlm.nih.gov> (last visited Feb. 18, 2003); among many others. There is no comprehensive list of all data centers in all areas of science and technology. However, there are over 100 federal data centers listed for global change research alone. See NASA's Global Change Master Directory, <http://gcmd.gsfc.nasa.gov> (last visited Feb. 13, 2003). It is also important to note that some of these data repositories, such as the NOAA National Data Centers, charge substantial fees for access. See *infra* notes 56-60 and accompanying text.

Scientific and technical articles, reports, and other information products generated by the federal government are also not copyrightable and available in the public domain. 17 U.S.C. § 105 (2000). Most research agencies have well-organized and extensive dissemination activities for such information, typically referred to as scientific and technical information, or "STI" (as distinct from data as such). These organizations include the National Library of Medicine, <http://www.nlm.nih.gov> (last visited Feb. 20, 2003); the National Agricultural Library, <http://www.nal.usda.gov> (last visited Feb. 18, 2003); the Defense Technical Information Center, <http://www.dtic.mil> (last visited Feb. 20, 2003); the Office of Scientific and Technical Information at the U.S. Department of Energy, <http://www.osti.gov> (last visited Feb. 18, 2003); and the NASA Scientific and Technical Information Program, <http://www.sti.nasa.gov> (last visited Feb. 20, 2003); among others.

Most of the STI products held by these repositories are available free of charge. Many agencies, however, also use the National Technical Information Service in the Department of Commerce, <http://www.ntis.gov/> (last visited Feb. 20, 2003), which makes additional STI available to the public for a fee. The Federal Depository Library Program provides yet another outlet for such information through its regional libraries, at http://www.access.gpo.gov/SU_docs/locators/findlibs (last visited Feb. 18, 2003). Finally, the National Archives and Records Administration provides permanent access to a subset of the STI resources, which it appraises and makes selectively available thirty years after their production, at <http://www.archives.gov/> (last visited Feb. 18, 2003).

46. David Banisar, Freedom of Information and Access to Government Records Around the World (July 2, 2002), at <http://www.freedominfo.org/survey.htm> (last visited Feb. 13, 2003) (indicating that over forty countries now have comprehensive laws to facilitate access to state records, and another thirty are in the process of enacting such statutes). For the situation in the European Union, see the European Commission's Green Paper, Public Sector Information: A Key Resource for Europe, *annexe 1* at 20-25 COM (1998) 585 [hereinafter Green Paper].

public-access policies are much less open than those applicable in the United States.⁴⁷ Notable examples of foreign sources of public domain data are the World Data Centers for geophysical, environmental, and space data⁴⁸ and the human genome databases in Europe and Japan.⁴⁹ However, a key issue for both the exploitation of public data resources and for cooperative research generally is the asymmetry between the United States and foreign government approaches to the public domain availability of scientific data.⁵⁰ This asymmetry has been deepened by the European Union's adoption in 1996 of an exclusive property right in non-copyrightable collections of data, as discussed in more detail in Part III.

2. Limiting Factors

Before delving into the growing pressures on the public domain, a number of countervailing policies and practices that limit free or open and unrestricted use of government-generated data and information must be noted. Important statutory exemptions to public domain access are based on national security concerns,⁵¹ the need to protect personal privacy of human subjects in research,⁵² and the need to respect confidential information.⁵³ These limitations on the public domain accessibility of federal government information, while often justified, must nonetheless be balanced against the rights and needs of citizens to access and use it.

Another limitation derives from the fact that government-generated data are not necessarily provided cost free. The federal policy for information dissemination, as set out in the Office of Management and Budget's ("OMB") Circular A-130, stipulates that such data should be made available at the marginal cost of dissemination—that is, the cost of fulfilling a user request.⁵⁴ That policy expressly excludes recouping the costs of production, much less making a profit.⁵⁵ In practice, the prices actually charged vary between marginal and

47. Weiss & Backlund, *supra* note 7, at 307; Yvette Plujimers & Peter Weiss, *Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts* (unpublished manuscript, on file with authors).

48. For a brief history of the World Data Center system and their general principles of operation, see the International Council for Science World Data Center System, at <http://www.ngdc.noaa.gov/wdc/> (last visited Feb. 13, 2003). See also links to the various World Data Center home pages, at <http://www.ngdc.noaa.gov/wdc/gdhomepg.html> (last visited Feb. 13, 2003).

49. See European Human Genome Database, available at <http://www.embl-heidelberg.de> (last visited Feb. 13, 2003); Human Genome Database of Japan, available at <http://www.ddbj.nig.ac.jp> (last visited Feb. 13, 2003).

50. The E.C. Directive on the legal protection of databases, *supra* note 17, does not prohibit proprietary rights in government-generated data (some governments assert these rights vigorously), nor does it mandate an exception for scientific uses of protected collections of data.

51. 5 U.S.C. § 552(b)(1) (2000).

52. 5 U.S.C. § 552(b)(6).

53. 5 U.S.C. § 552(b)(4).

54. Office of Management and Budget Circular A-130, § 8a(7) ("Information Management Policy—Avoiding Improperly Restrictive Practices") (Feb. 8, 1996), available at <http://www.whitehouse.gov/omb/circulars/a130/a130.html>.

55. *Id.* at Appendix § 4.

incremental cost pricing arrangements.⁵⁶ Charges higher than the marginal cost of dissemination can create substantial barriers to access, particularly for academic research, which may require the use of large portions or the entire contents of huge databases for modeling or data mining applications. Nevertheless, this policy differs from that of most other countries, in which governmental or quasi-governmental agencies may exploit public information at full cost recovery rates and may also invoke the protection of that information under intellectual property law.⁵⁷

In practice, a major barrier to accessing government-generated data and information arises from the failure of agencies to disseminate or preserve them for long-term availability. As a result, a massive amount of public domain information is either hidden from public view or is irretrievably lost.⁵⁸ Although this generally tends to be less problematic for scientific information, the well-known problems of inadequate documentation and organization are particularly acute for some areas of scientific data and information, particularly in the life sciences.⁵⁹ When the problem is a simple failure to disseminate, the Freedom of Information Act ("FOIA") may sometimes provide an antidote.⁶⁰ However, the process of filing a FOIA request is time consuming and bureaucratic.

Another limitation derives from OMB Circular A-76, which bars the government from directly competing with the private sector in providing information products and services.⁶¹ This policy substantially narrows the amount and type of information that the government can undertake to produce and disseminate. As regards science, the well-accepted view is that basic research, together with its supporting data, constitute public goods that properly fall within the sphere of government activity, although the boundary between what is considered appropriate public or private functions continues to shift in this and other respects.

Finally, the government is required to respect the proprietary rights in data and information originating from the private sector that are made available for government use or, more generally, for regulatory and other purposes, unless

56. In some exceptional circumstances, agencies may charge full incremental cost recovery prices. For example, 15 U.S.C. § 1534 authorizes the Secretary of Commerce to charge fair market value for the data from the NOAA National Data Centers, although lower prices are allowed to be charged to educational organizations.

57. For the pricing policies for government data in the European Union, see Green Paper, *supra* note 46. See also PIRA INTERNATIONAL, COMMERCIAL EXPLOITATION OF EUROPE'S PUBLIC SECTOR INFORMATION, FINAL REPORT FOR THE EUROPEAN COMMISSION, DIRECTORATE GENERAL FOR THE INFORMATION SOCIETY (2000).

58. See generally 1 U.S. NATIONAL COMMISSION ON LIBRARIES AND INFORMATION SCIENCE, A COMPREHENSIVE ASSESSMENT OF PUBLIC INFORMATION DISSEMINATION, FINAL REPORT (2001). In the United States, some state governments also seek copyright protection of model codes and other quasi-legislative materials. See, e.g., *Veek v. S. Bldg. Code Cong. Int'l*, 293 F.3d 791 (5th Cir.), *cert. granted*, 123 S. Ct. 650 (2002).

59. See BITS OF POWER, *supra* note 1, at 70-74.

60. Freedom of Information Act (FOIA), 5 U.S.C. § 552 (2002).

61. Office of Management and Budget Circular A-76 (Aug. 8, 1983, rev. 1999), Performance of Commercial Activities, § 5(c), available at <http://www.whitehouse.gov/omb/circulars/a076/a076.html> (last visited Feb. 13, 2003).

expressly exempted.⁶² To the extent that more of the production and dissemination functions of research data are shifted from the public to the private sector, this limitation becomes more potent. Moreover, this trend easily gives way to pressures to prevent reasonable commercial uses of data that firms must submit to government for regulatory purposes and can even result in back door, de facto intellectual property rights that protect scientific data.⁶³ By the same token, foreign governments that opt to protect and commercialize their own data may seek to restrict their open dissemination and use by the U.S. government for research and other public-interest purposes.⁶⁴

B. Government-Funded Data: Between the Research Commons and Commercial Applications

A second major source of public domain data and information for scientific research is that which is produced in academic or other not-for-profit institutions with government and philanthropic funding. However, databases and other information goods produced in these settings become presumptively protectible under any relevant intellectual property regime unless affirmative steps are taken to place such material in the public domain. In this case, the public domain must be actively created, rather than passively conferred.

This component of the public domain results from the contractual requirements of the granting agencies in combination with long-standing norms of science. These norms aspire to implement “full and open access” to scientific data as well as the sharing of research results so as to promote new endeavors. The policy of full and open access or exchange has been defined in various U.S. government policy documents and in National Research Council (“NRC”) reports in the following terms: “[D]ata and information derived from publicly funded research are [to be] made available with as few restrictions as possible, on a nondiscriminatory basis, for no more than the cost of reproduction and distribu-

62. See, e.g., TREVOR M. COOK, *THE PROTECTION OF REGULATORY DATA IN PHARMACEUTICAL AND OTHER SECTORS*, Preface, 7 (2000) (stressing that concerns about the confidentiality of regulatory data, including the results of clinical trials, have mainly surfaced in the past twenty-five years). Since 1982, the United States has adopted provisions to protect regulatory data submitted to federal agencies in connection with pesticides, and it has imposed regulatory exclusivity provisions for medical data since 1984. *Id.* at 4-01. These provisions reportedly “provide a *de facto* measure of . . . data protection” to new chemical entities for five years and they give three years of protection for “data filed . . . in support of . . . chemical entities which have already been approved for use in medicines but [for] which fresh authorizations are [to be] based on new clinical investigations.” *Id.* For analogous provisions that may confer even longer periods of protection in the European Union, Australia, and New Zealand, see *id.* at Preface 6-7, 3-01, 5-01.

63. See, e.g., *Ruckelshaus v. Monsanto Co.*, 467 U.S. 986, 1019-20 (1984); *Bayer, Inc. v. Canada* (Attorney General), [1999] F.C.A.D.J. 142. Some obligations in this regard have even been codified as international minimum standards under article 39.3 of the TRIPS Agreement. Agreement on Trade-Related Aspects of Intellectual Property Rights, April 15, 1994, 33 I.L.M. 81 [hereinafter TRIPS Agreement]. See generally Carlos Correa, *Public Health and International Law: Unfair Competition Under the TRIPS Agreement Article 39.3: Protection of Data Submitted for Registration of Pharmaceuticals*, 3 CHI. J. INT’L L. 69 (2002).

64. See *infra* notes 543-50 and accompanying text.

tion”⁶⁵ (that is, the marginal cost, which on the Internet is zero). This policy is promoted by different U.S. government agencies with varying degrees of success. It also applies to most government-funded or cooperative research arrangements, particularly in large, institutionalized research programs, such as “global change” studies or the human genome project, and even to smaller-scale collaborations involving individual investigators who are not otherwise affiliated with private-sector partners.⁶⁶

Because government agencies are funding these efforts, they are in a position to reinforce the underlying norms of science by suitable contractual provisions that regulate access to data before and after publication of the research results. Basic research grants in particular attempt to ensure that government-funded data enter the upstream processes of scientific research as an input available from the public domain.⁶⁷

In addition, the research articles published in scientific journals will themselves become subject to the balancing of public and private interests that occurs under the prevailing intellectual property regime. For example, traditional copyright principles have consigned facts, ideas, data, and findings to the public domain; there are codified exceptions for research and educational uses, and exceptions for fair uses or private use to advance research and other public-interest goals have been granted.⁶⁸

Scientists also hold much data in their individual capacities, which will be made available on a transactional basis. In principle, the government science agencies’ contractual specifications that promote eventual access to government-funded data reinforce long-standing norms of science, which are traditionally premised on open access and the sharing ethos.⁶⁹ While the real-world implementation of these norms has always been imperfect and is now subject to intense countervailing pressures that are described below, one must nonetheless emphasize, at the outset, that the government’s baseline default rules and the norms of science remain mutually reinforcing with respect to data as an input into scientific research. This convergence traditionally makes it harder for scientists to violate these contractual and cultural norms with regard to access to government-funded data as upstream inputs for research purposes.

Even so, scientific data have a dual nature in the sense that they are also outputs of the scientific process, and these outputs—suitably aggregated—become inputs once again into the national system of innovation. Here,

65. BITS OF POWER, *supra* note 1, at 1.

66. See FULL AND OPEN EXCHANGE, *supra* note 8 at 2; NATIONAL SCIENCE FOUNDATION, GRANT GENERAL CONDITIONS (GC-1) #36 (2001).

67. See, e.g., NATIONAL SCIENCE FOUNDATION 95-26, GRANT POLICY MANUAL § 734 (1995) [hereinafter GRANT POLICY MANUAL]; see also *infra* note 72.

68. See *supra* note 12; Thomas Dreier, *Balancing Proprietary and Public Domain Interests: Inside or Outside Proprietary Rights*, in EXPANDING THE BOUNDARIES OF IP, *supra* note 13, at 295, 301, 303-09. See generally Jaap H. Spoor, *General Aspects of Exceptions and Limitations to Copyright: General Report*, in THE BOUNDARIES OF COPYRIGHT—ITS PROPER LIMITATIONS AND EXCEPTIONS 27-41 (Libby Baulch et al. eds., 1997) (providing the most recent comparative survey of existing law).

69. See *supra* note 8.

however, very different government policies may apply. The non-proprietary default rule built into government-funded data exists side-by-side with a second, posterior set of default rules that encourage research entities outside the federal system to transfer government-funded research from the public to the private sector, usually by means of intellectual property rights.⁷⁰

It becomes necessary to look beyond the surface to gain some deeper insights into how the stated policies and rules built around government-funded data are actually implemented in academic science. In this regard, it is useful to further subdivide the producers of government-funded data into two distinct but partially overlapping categories.

In one category, the research takes place in a highly structured academic setting, with relatively clear rules set by government funding agencies that determine the rights of researchers in the production, dissemination, and use of data. Publication of the research results constitutes the primary organizing principle. Traditionally, the rules and norms that apply in this sphere, which we call the zone of formally regulated access to data, aim with varying degrees of success to achieve a bright line of demarcation between the public and private rights to the data being generated.

In the other category, individual scientists establish their own interpersonal relationships and networks with other colleagues, largely within their specialized research communities. In this category, which we call the zone of informal data exchanges, scientists may generate and hold their data subject to their own interests and to the interplay of norms, rules, and competitive strategies that may deviate considerably from the practices established within the zone of formal exchanges.

1. The Zone of Formally Regulated Access to Data

As previously noted, when federal government agencies fund research projects that include the production of scientific data in not-for-profit academic institutions, they typically stipulate a set of rights and obligations binding the agency and the principal investigator with regard to those data. For example, the grant may entitle the investigator to a period of exclusive use of the relevant data prior to publication of the research results. It may also mandate that the data sets collected under the grant become freely available for others to use following the publication of results or upon the expiration of the exclusive use period, and to this end, it may further specify that such materials should be deposited in government or university data centers or archives.⁷¹

70. See, e.g., Rai, *supra* note 33, at 95-115; see also Brett Frischmann, *Innovation and Institutions: Rethinking the Economics of U.S. Science and Technology Policy*, 24 VT. L. REV. 347, 353, 395-413 (2000).

71. See, e.g., National Science Foundation Office of Polar Programs, *Guidelines and Award Conditions for Scientific Data* (1998); National Aeronautics and Space Administration, *Science Policy Guide* (1996); National Science Foundation Division of Ocean Sciences 94-126, *Policy for Oceanographic Data* (1994).

a. Contractually Reinforcing the Sharing Ethos

Contractual requirements vary by agency and discipline and even by research programs undertaken within the same agency. Different agencies may address these issues with more or less attentiveness, and different offices within the same agency may have diverse requirements as well.⁷²

72. Although a comprehensive assessment of specific data rights across all federal science agency grants and contracts is beyond the scope of this discussion, a brief review of some of the most common provisions is instructive.

For example, the standard clause on “Dissemination and Sharing of Research Results” in a National Science Foundation (“NSF”) grant provides as follows:

- a. Investigators are expected to promptly prepare and submit for publication, with authorship that accurately reflects the contributions of those involved, all significant findings from work conducted under NSF grants. Grantees are expected to permit and encourage such publication by those actually performing that work, unless a grantee intends to publish or disseminate such findings itself.
- b. Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. . . .
- c. Investigators and grantees are encouraged to share software and inventions created under the grant or otherwise make them or their products widely available and usable.
- d. The NSF normally allows grantees to retain principal legal rights to intellectual property developed under NSF grants to provide incentives for development and dissemination of inventions, software and publications that can enhance their usefulness, accessibility and upkeep. Such incentives do not, however, reduce the responsibility that investigators and organizations have as members of the scientific and engineering community to make results, data and collections available to other researchers.

GRANT POLICY MANUAL, *supra* note 67, at § 734.

The National Institutes of Health (“NIH”) is currently developing a statement on sharing research data that:

[E]xpects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers. Investigators submitting an NIH application will be required to include a plan for data-sharing or to state why data-sharing is not possible. This statement will apply to extramural scientists seeking grants, cooperative agreements, and contracts as well as intramural investigators.

Release, NIH Announces Draft Statement on Sharing Research Data, (March 1, 2002) (NOTICE: NOT-OD-02-035), *available at* <http://grants1.nih.gov/grants/guide/notice-files/NOT-OD-02-035.html> (last visited Jan. 10, 2003).

The announcement goes on to say:

There are many reasons to share data from NIH-supported studies. Sharing data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, and permits the creation of new data sets when data from multiple sources are combined. By avoiding the duplication of expensive data collection activities, the NIH is able to support more investigators that it could if similar data had to be collected *de novo* by each applicant.

Id.

Similarly, NASA’s Data Availability Policy states that:

[R]eady access to data from NASA research programs and missions (via modern data archiving and communications technologies) by researchers not directly involved in the program increases the return on NASA research investments. It is therefore NASA policy that non-proprietary scientific data obtained from NASA programs and missions will be made publicly available in usable form as quickly as possible.

In general, the common thrust of these different types of clauses is to ensure that the data collected or generated by grantees will be openly shared with other researchers, at least following some specified period of exclusive use—typically ranging from six to twenty-four months—or until the time of publication of the research results based on those data.⁷³ This relatively brief period is intended to give the grantee sufficient time to produce, organize, document, verify, and analyze the data being used in preparation of a research article or report for scholarly publication. In many cases, the data are placed in a public archive upon publication, or at the expiration of the specified period of exclusive use and are expressly designated as free from legal protection, or they are expected to be made available directly by the researcher to anyone who requests access.

In most cases, publication of research results marks the point at which data produced by government-funded investigators should become generally available. The standard research grant requirement or norm has been that once publication occurs, it will trigger public disclosure of the supporting data. To the extent that this requirement is implemented in practice, it represents the culmination of the scientific norms of sharing.⁷⁴ From this point forward, access to the investigator's results will depend on the method chosen to make the underlying data publicly available and on the traditional legal norms—especially those of copyright law—that govern scientific works.

These organizing principles derive historically from the premise that academic researchers typically are not driven by the same motivations as their counterparts in industry. Public-interest research is not dependent on the maximization of profits and value to shareholders through the protection of proprietary rights in information. Rather, the motivations of not-for-profit scientists are predominantly rooted in intellectual curiosity, the desire to create new knowledge, peer recognition and career advancement, and the promotion of the public interest.⁷⁵ As R. Stephen Berry, the Home Secretary for the National Academy of Sciences, recently noted:

Scientists are not, for the most part, motivated to do research to make money. If they were, they would be in different fields. The primary motivation for most research scientists is the desire for influence and impact on the thinking of others about the natu-

BITS OF POWER, *supra* note 1, at 80 (quoting NASA Science Policy Guide (1996)). The policy goes on to provide a list of competing factors that need to be considered in determining data rights, and presents examples of data rights that have been used, mostly variations on the length of the initial period of an investigator's proprietary use.

73. See BITS OF POWER, *supra* note 1, at 79. One of the recommendations of that report was that all scientists conducting publicly funded research should make their data available immediately, or following a reasonable period of time for proprietary use. *Id.* at 11.

74. See, e.g., NATIONAL RESEARCH COUNCIL, COMMUNITY STANDARDS FOR SHARING PUBLICATION-RELATED DATA AND MATERIALS (2002) [hereinafter COMMUNITY STANDARDS] (discussing these norms and requirements in the biological sciences).

75. See, e.g., Rebecca Eisenberg, *Proprietary Rights and the Norms of Science in Biotechnology Research*, 97 YALE L.J. 177, 178 (1987) [hereinafter Eisenberg, *Proprietary Rights*] (“The scientific community rewards those who make original contributions to the common stock of knowledge by giving them professional recognition.”).

ral world—unless the desire for their own personal understanding is even stronger. . . . The currency of the researcher is the extent to which her or his ideas influence the thinking of others. . . . What this implies is that the distribution of the results of research has an extremely high priority for any working scientists, apart from those whose work is behind proprietary walls.⁶

Science policy in the United States has long taken for granted that these values and goals are best served by the maximum availability and distribution of research results, at the lowest possible cost, with the fewest restrictions on use, and the promotion of reuse and integration of the fruits of existing results in new research. The placement of scientific and technical (“S&T”) data and databases in the public domain, and the established policy of full and open access to such resources in the government and academic sectors⁷⁷ reflect these values and serve these goals.

b. Legal Rules that Support the Sharing Ethos

Assuming that the relatively standard grant rules apply and that scientific data are fully disclosed at the time of publication in the manner previously described, the intellectual property norms that traditionally came into play in the post-publication phase were, on the whole, consonant with the cultural norms of the scientific enterprise. First, data as such were not generally considered subject matter eligible for copyright protection in most jurisdictions, and the few deviant jurisdictions that thought otherwise were overruled by the 1991 Supreme Court decision in *Feist Publications, Inc. v. Rural Telephone Service Co.*⁷⁸ Furthermore, the United States has not yet adopted a *sui generis* legal regime to protect non-copyrightable collections of information as the European Union did in 1996,⁷⁹ while most claims arising under unfair competition law are forfeited with voluntary disclosure to the public.⁸⁰

Second, the power of contracts to regulate the dissemination of publicly disclosed data was inherently weak in the pre-digital age owing to the inability of the contracting parties to sue third parties who obtained the relevant data outside of the contractual relationship.⁸¹ It was the inherent weakness of these

76. R. Stephen Berry, *Is Electronic Publishing Being Used in the Best Interests of Science? The Scientists' View*, 2001 INT'L J. MOLECULAR SCI. 133, 134 (2001).

77. See *supra* notes 65, 72 and accompanying text.

78. 499 U.S. 340 (1991) (holding the factual information in the white pages of a telephone book lacked creativity and originality in selection and arrangement, and was not copyrightable). *But see* Paula Baron, *Back to the Future: Learning from the Past in the Database Debate*, 62 OHIO ST. L. J. 874 (2001); Robert C. Denicola, *Copyright in Collections of Facts: A Theory for the Protection of Nonfiction Literary Works*, 81 COLUM. L. REV. 516, 528, 539-40 (1981) (stressing need for compiler's incentives); Jane C. Ginsburg, *Creation and Commercial Value: Copyright Protection of Works of Information*, 90 COLUM. L. REV. 1865 (1990) [hereinafter Ginsburg, *Commercial Value*].

79. E.C. Database Directive, *supra* note 17.

80. See generally RESTATEMENT (THIRD) OF UNFAIR COMPETITION §§ 38-45 (1995) (allowing claim of misappropriation of trade secrets, but not recognizing any broader misappropriation claim rooted in copying as such). However, state law doctrines of misappropriation may nonetheless apply to wholesale duplication of databases to a still unknown extent. See, e.g., *Int'l News Serv. v. Assoc. Press*, 248 U.S. 215 (1918); *Nat'l Basketball Ass'n v. Motorola, Inc.*, 105 F.3d 841 (2d Cir. 1996).

81. See PAUL GOLDSTEIN, *COPYRIGHT'S HIGHWAY: FROM GUTENBERG TO THE CELESTIAL JUKEBOX* 27 (1994) (noting the lack of need for copyright protection before the printing press).

“two-party deals” to regulate generally the dissemination of intangible literary and artistic productions that gave rise to copyright and neighboring rights laws,⁸² which impose an unbargained-for set of default rules upon all those who gain access to such productions.⁸³ If anything, the traditional use of grants and contracts in the domain of government-funded data was to implement the sharing norms of science, as shown above.⁸⁴

Third, with regard to the data and information contained in published scientific works, the balance of public and private interests traditionally struck by U.S. copyright law was particularly favorable to second-comers and researchers in general. Collections of information distributed in hard copies, such as directories, handbooks, and other useful compilations of facts or data, are copyrightable only to the extent that they manifest a minimum quantum of original and creative authorship.⁸⁵ Typically, the requisite degree of authorship is revealed in the compiler’s criteria for selecting, arranging, and documenting the data assembled in any given compilation.⁸⁶ However, the facts and data contained in a copyrightable scientific work, for example, are ineligible for protection as are any “idea, procedure, process, system, method of operation, concept, principle or discovery” it contains.⁸⁷

Moreover, after passing the threshold of copyright law, even those “factual works” that do manifest the minimum degree of creative authorship are likely to obtain only a “thin” scope of protection at the infringement stage.⁸⁸ Because the facts and data in scientific works are not copyrightable subject matter, and only the creative selection or arrangement is protectible, a second-comer can, in principle, borrow the first-comer’s disparate data while varying the organizational format. As the Supreme Court noted in *Feist*, the traditional copyright approach to scientific and other factual works thus strikes a balance between incentives to invest and free competition that tends to err on the side of second-comers.⁸⁹ In effect, by severely curtailing the first-comer’s ability to control follow-on applications of factual content,⁹⁰ copyright law in this area operates as a kind of roving unfair competition law that protects the authors of scientific works mainly against wholesale duplication. In the United States, these limitations are thought to have constitutional underpinnings, in keeping with First

82. See Reichman & Franklin, *supra* note 16, at 897-99.

83. See, e.g., Peter A. Jaszi, *Goodbye to All That—A Reluctant (and Perhaps Premature) Adieu to a Constitutionally—Grounded Discourse of Public-Interest in Copyright Law*, 29 VAND. J. TRANSNAT’L L. 595, 599-600 (1996) (stressing economic and cultural bargain between authors and users).

84. See *supra* note 72 (citing specific examples).

85. 17 U.S.C. §101 (2000) (definition of compilations); § 102(a)-(b); § 103; *Feist Publ’ns, Inc. v. Rural Tel. Servs. Co.*, 499 U.S. 340, 345 (1991).

86. See *Feist*, 499 U.S. at 348; see also *Warren Publ’g Inc. v. Microdos Data Corp.*, 52 F.3d 950, 956 (11th Cir. 1995); *Bellsouth Adver. & Publ’g Corp. v. Donnelly Info. Publ’g Inc.*, 999 F.2d 1436, 1446 (11th Cir. 1993) (en banc); *Key Publ’ns, Inc. v. Chinatown Publ’g Enter., Inc.*, 945 F.2d. 509, 514 (2d Cir. 1991).

87. 17 U.S.C. § 102(b); *Harper & Row Publ’rs, Inc. v. Nation Enters.*, 471 U.S. 539 (1985).

88. *Feist*, 499 U.S. at 349.

89. *Id.* at 349-50.

90. Notwithstanding 17 U.S.C. § 103 and § 106(2), which the Supreme Court limited in this respect.

Amendment rights protecting freedom of speech and with the role of a robust public domain in democratic discourse.⁹¹

One of the largest categories of scientific information in the United States thus consists of ineligible collections of data or the non-copyrightable contents of otherwise copyrightable works, including databases, articles, or reference books. This category of public domain information, while highly distributed among all types of proprietary works, plays a fundamental role in supporting research and education, especially in the data-intensive sciences. However, strenuous efforts are being made to devise new forms of protection for all of this previously unprotectible subject matter, as explained in Part III of this article.

When scientific works, including compilations of data and information, do attract copyright protection, the copyright laws of most countries contain codified limitations, exceptions, or immunities that favor teaching, research, and other educational activities.⁹² In European Union countries, there is a “private use” exception, which to some extent overlaps the much broader “fair use” exception in U.S. copyright law.⁹³ Also, in the European Union compulsory licenses may be enacted to promote public policy goals, including research and education, without depriving authors of an economic return from these uses of their works.⁹⁴ For example, some countries allow the photocopying of copyrighted journals for research purposes only in exchange for compensatory payments that are increasingly worked out with collection societies representing authors’ interests in these matters.⁹⁵

In U.S. copyright law, considerable emphasis has traditionally been placed on the fair use exception to copyright protection,⁹⁶ which permits certain uses to

91. *Harper & Row*, 471 U.S. at 582; *see also* 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 (2000) [hereinafter Benkler, *Constitutional Bounds*]; Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 N.Y.U. L. REV. 354 (1999) [hereinafter Benkler, *Free as the Air*]; James Boyle, *Foucault in Cyberspace: Surveillance, Sovereignty, and Hardwired Censors*, 66 U. CIN. L. REV. 177 (1997); Marci A. Hamilton, *A Response to Professor Benkler*, 15 BERKELEY TECH. L. J. 605 (2000); Neil Netanel, *Locating Copyright Within the First Amendment Skein*, 54 STAN. L. REV. 1 (2001).

92. *See, e.g.*, Dreier, *supra* note 68; Sam Ricketson, *International Conventions and Treaties*, in BOUNDARIES OF COPYRIGHT, *supra* note 68, at 3, 5-10 (stressing recurring exceptions in national copyright laws for private study, and for “use for scientific and research purposes,” in addition to provisions allowing use for teaching purposes); *see also* Ruth Okediji, *Toward an International Fair Use Doctrine*, 39 COLUM. J. TRANSNAT’L L. 75 (2000).

93. *See* Adolf Dietz, *Germany*, in BOUNDARIES OF COPYRIGHT, *supra* note 68, at 265, 269 (noting rights of free reproduction or other private use, sometimes subject to an obligation to remunerate, under articles 53-54(a) of German copyright law); Yves Gaubiac, *France*, in *id.* at 226, 231 (noting exception for private noncommercial use to promote private study and research under French law).

94. *See, e.g.*, Dreier, *supra* note 68; Ricketson, *supra* note 92, at 9, 14; Spoor, *supra* note 68.

95. *See, e.g.*, LUCIE M.C.R. GUIBAULT, COPYRIGHT LIMITATIONS AND CONTRACTS: AN ANALYSIS OF THE CONTRACTUAL OVERRIDABILITY OF LIMITATIONS ON COPYRIGHT 81-82 (2002) (discussing mandatory collective administration of reprography right under French copyright law); Dietz, *supra* note 93, at 269 (stressing basic permissibility of private copying and reprography even when subject to collective agreements on equitable remuneration).

96. 17 U.S.C. § 107 (2000).

be made of otherwise protected content under limited circumstances, in order to advance the public interest in certain privileged policy goals without unduly burdening the authors' incentive to create.⁹⁷ On a case-by-case basis, this doctrine may permit uses of otherwise protected expression,⁹⁸ especially for such purposes as illustration, teaching, research, verification, and news reporting.⁹⁹ However, the strength of this exception varies with judicial attitudes and from period to period, and its consistency with international intellectual property law has been called into question.¹⁰⁰

Because many so-called fair uses are allowed only in the context of not-for-profit research or education, this category of "public domain uses," though relatively small, is especially important in the research context. While commentators have not typically associated fair uses and other exceptions with the "public domain,"¹⁰¹ a number of traditionally practiced immunities and exceptions, including fair use, may be construed as functional equivalents of public domain uses, especially where science and education are concerned.¹⁰²

Applications of the fair use exception tend to be controversial, however, even with regard to educational and research activities, where publishers often emphasize the harm to markets resulting from ad hoc judicial concessions to users of copyrighted materials in particular situations.¹⁰³ Recently, the federal appellate courts have tended to restrict the fair use exception when technical means to overcome market failure are shown to exist,¹⁰⁴ and the Digital Millennium Copyright Act ("DMCA") has severely cut back on the application of fair use to online transmissions.¹⁰⁵ At the same time, there is growing interest in doctrines of misuse of copyrights, which courts have used to strike down

97. See, e.g., Julie E. Cohen, *Lochner in Cyberspace: The New Economic Orthodoxy of Rights Management*, 97 MICH. L. REV. 462, 468-80 (1998); William W. Fisher, III, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1661 (1998); Wendy J. Gordon, *Fair Use as Market Failure: A Structural and Economic Analysis of the Betamax Case and Its Predecessors*, 82 COLUM. L. REV. 1600 (1982).

98. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994); *SunTrust Bank v. Houghton Mifflin Co.*, 268 F.3d 1257 (11th Cir. 2001) (finding at preliminary injunction stage, publisher of *The Wind Done Gone* was entitled to fair-use defense against copyright infringement claim); *Williams & Wilkins Co. v. United States*, 487 F.2d 1345 (Ct. Cl. 1973).

99. See 17 U.S.C. § 107 (preambular uses). See generally PAUL GOLDSTEIN, *COPYRIGHT: PRINCIPLES, LAW AND PRACTICE* § 10.1.5 (1996).

100. See, e.g., Okediji, *supra* note 92.

101. See, e.g., GOLDSTEIN, *supra* note 99, at §10.1 (noting theories of fair use sounding in "privilege," "equitable rule of reason," "market failure," and "public benefit" theories).

102. Cf., e.g., *id.* at §10.2.1 (stressing cultural and social values of an educated public that follow from preambular fair uses favoring teaching, scholarship, and research).

103. 17 U.S.C. § 107(4).

104. *Princeton Univ. Press v. Michigan Document Servs.*, 99 F.3d 1381 (6th Cir. 1996); *Am. Geophysical Union v. Texaco, Inc.*, 60 F.3d 913 (2d Cir. 1994); see also Gordon, *supra* note 97.

105. 17 U.S.C. § 1201; David Nimmer, *A Riff on Fair Use in the Digital Millennium Copyright Act*, 148 U. PA. L. REV. 673 (2000); Pamela Samuelson, *Mapping the Digital Public Domain*, 66 LAW & CONTEMP. PROBS. 147 (Winter/Spring 2003); see also Ginsburg, *U.S. Initiatives*, *supra* note 18, at 62-67; Maureen A. O'Rourke, *Copyright Preemption After the Pro-CD Case: A Market Based Approach*, 12 BERKELEY TECH. L.J. 53 (1997).

licensing restrictions that unreasonably extend the statutory bundle of exclusive rights or that otherwise impose unreasonable restraints on trade.¹⁰⁶

Finally, it should be recalled that copyrights eventually expire when the statutory period of protection ends, and the public then becomes the residual owner of all previously protected works. The basic term of copyright protection is very long, especially in the United States and the European Union, where it now lasts for the life of the author plus seventy years.¹⁰⁷ The United States separately protects corporate works (technically, works made for hire) for either a term of ninety-five years from first publication or 120 years from creation, whichever is shorter.¹⁰⁸ The relevant international minimum standard requires a term of life plus fifty years for works by human authors and at least fifty years of protection for works by corporate authors.¹⁰⁹

Works whose copyrights have expired constitute an enormous body of freely available literature and information with great cultural and historical significance, and some materials in this category have obvious relevance to certain types of research, especially in the social sciences and the humanities. Even some of the “hard” sciences can derive substantial value from public domain data and information that are decades or even many centuries old. For example, the extraction of environmental information from a broad range of historical sources can help to establish climatological trends, or assist in identifying or better understanding a broad range of natural phenomena.¹¹⁰ Ancient Chinese writings are expected to be useful in identifying herbal medicines for modern pharmaceutical development.¹¹¹ Proposals for more systematic and ambitious databases concerning traditional know-how and medicines are on the table,¹¹² although these proposals would almost invariably subject such information to new forms of intellectual property rights that would limit their availability in the public domain.¹¹³ On the whole, because of the long lag time before entering the

106. See, e.g., Dreier, *supra* note 68, at 311-12; Brett Frischmann & Dan Moylan, *The Evolving Common Law Doctrine of Copyright Misuse: A Unified Theory and Its Application to Software*, 15 BERKELEY TECH. L.J. 865 (2000).

107. 17 U.S.C. §§ 302, 303 (2000); Council Directive 93/98/EEC of 29 October 1993 Harmonizing the Term of Protection of Copyright and Certain Related Rights, 1993 O.J. (L 290) 9; see also *Eldred v. Reno*, 239 F.3d 372 (D.C. Cir. 2001), *cert. granted sub nom*, *Eldred v. Ashcroft*, 534 U.S. 1062, and *cert. amended*, 534 U.S. 1160 (2002).

108. 17 U.S.C. § 302(c).

109. TRIPS Agreement, *supra* note 63, art. 12; Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886 (as last revised July 24, 1971, and amended Oct. 2, 1979), 828 U.N.T.S. 221, art. 7(1) [hereinafter Berne Convention]. See generally J. H. Reichman, *The Duration of Copyright and the Limits of Cultural Policy*, 14 CARDOZO ARTS & ENT. L.J. 625 (1996).

110. FULL AND OPEN EXCHANGE, *supra* note 8, at 3.

111. BITS OF POWER, *supra* note 1, at 73.

112. See, e.g., Graham Dutfield, *TRIPS-Related Aspects of Traditional Knowledge*, 33 CASE W. RES. J. INT'L L. 239 (2001); Traditional Knowledge, Intellectual Property and Indigenous Culture, Symposium presented at the Benjamin N. Cardozo School of Law (Feb. 21-22, 2002).

113. See, e.g., William van Caenegem, *The Public Domain: Scientia Nullius?*, 2002 E.I.P.R. 324, 325, 328-30 (2002) (warning that some constructions of the public domain may be used to dispossess rights of indigenous peoples deriving from different social and cultural constructs).

public domain, most of the information in this category of lapsed protection lacks relevance to most types of state-of-the-art research.

c. Countervailing Policies: The Commercialization of Government-Funded Research

In the academic sector, the predominant norms remain those of open disclosure and the sharing of research data at the time of publication, if not before, and in many cases, this ethos also entails the placement of the data derived from federally funded research in public data centers and archives. Nevertheless, diverse policy incentives and economic pressures have increasingly induced research universities and academics to protect or commercialize their research data, rather than place them in the public domain. The costs and tuitions of higher education have far outpaced inflation, so there are direct economic needs for the universities to generate new sources of income wherever possible.¹¹⁴ Perhaps most significant, the 1980 Bayh-Dole Act¹¹⁵ has encouraged academics to protect and commercialize the fruits of their federally funded research,¹¹⁶ especially in the potentially lucrative biomedical research area.¹¹⁷

While Bayh-Dole technically applies only to data products otherwise eligible for patent protection,¹¹⁸ it supports the broader principle that universities should seek to commercialize applications of government-funded research products in general. Under Bayh-Dole, universities have moved away from policies that favor pure research, both for its own sake and as a tool for advancing higher education. As the costs of education skyrocket, and government funding fails to keep up in many areas, universities have aggressively sought to exploit commercial applications of research results, with an eye toward maximizing returns on investment.¹¹⁹

There is evidence that the Bayh-Dole Act has exerted a positive effect on technological innovation and that it has generated fruitful public-private part-

114. Regarding the steep rises in higher education costs, see NATIONAL CENTER FOR PUBLIC POLICY AND HIGHER EDUCATION, *LOSING GROUND: A NATIONAL STATUS REPORT ON THE AFFORDABILITY OF AMERICAN HIGHER EDUCATION* (2002), available at http://www.highereducation.org/reports/losing_ground/ar.shtml/ (last visited Jan. 10, 2003).

115. Pub. L. No. 96-517, § 6(a), 94 Stat. 3015, 3019-28 (1980) (codified as amended at 35 U.S.C. §§ 200-212 (2000)).

116. "It is the policy and objective of Congress to use the patent system to promote the utilization of inventions arising from federally supported research or development . . . [and] to promote the collaboration between commercial concerns and nonprofit organizations, including universities . . ." 35 U.S.C. § 200 (2000).

117. Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 *LAW & CONTEMP. PROBS.* 289 (Winter/Spring 2003); see also Frischmann, *supra* note 70, at 397-413; Rai, *supra* note 33, at 95-100.

118. 35 U.S.C. § 202(a) (2000).

119. Note, however, that generating revenue for universities was not the goal of the Bayh-Dole Act. Rai & Eisenberg, *supra* note 117, at 300. Technically, the Bayh-Dole Act requires that the profits accruing to the beneficiary nonprofit organizations "be utilized for the support of scientific research or education." See 35 U.S.C. § 202(c)(7) (2000).

nerships for commercial exploitation of academic research.¹²⁰ At the same time, the policies promoting downstream application of university research results under Bayh-Dole have increasingly come into conflict with the policies favoring full and open access to research data and with the larger educational and public interest mission of universities.¹²¹

First of all, because Bayh-Dole inclines university administrations and individual academics to seek opportunities for the commercial exploitation of federally funded research, they are tempted to hoard data, to refrain from divulging them fully, and to conduct research operations under rules of secrecy and confidentiality. In doing so, they are increasingly likely to treat data collections as private goods in support of commercial exploitation of related applications under restrictive licensing agreements.¹²²

A second development stemming in part from the blurring and collapse of the boundaries between basic and applied research, especially in biotechnology, is the discovery of ways to commercialize upstream aggregates of data as research tools and products.¹²³ This development greatly augments the potential value of some scientific databases, and raises obstacles to the dissemination and use of research data as further inputs into the research process that were seldom previously encountered.¹²⁴ In this environment, the value of the data supporting a patent application under Bayh-Dole may ultimately exceed the value of the invention being claimed, and more and more patent applications cover electronic information tools built around aggregates of data.

Similarly, the data supporting published scientific research may have potential commercial value to the originating institution and the responsible team of

120. Rebecca S. Eisenberg, *Public Research and Private Development: Patents and Technology Transfer in Government-Sponsored Research*, 82 VA. L. REV. 1663 (1996) [hereinafter Eisenberg, *Public Research*].

121. See, e.g., Eisenberg, *Proprietary Rights*, *supra* note 75, at 180 (finding that although the patent system and the norms of science have much in common, “the conjunction may nonetheless cause delay in the dissemination of new knowledge and aggravate inherent conflict between the norms and reward structure of science”); Rebecca S. Eisenberg & Richard Nelson, *Public vs. Proprietary Science: A Fruitful Tension*, DAEDALUS, Spring 2002, at 92 (“Even if expected practical benefits make patentable outcomes likely and motivate private firms to pay for the research, public funding might still be justified to increase the open domain of commonly owned knowledge upon which scientists may draw freely in future research.”); Rai, *supra* note 33, at 109-37 (stressing negative impact on public domain); see also Avital Bar-Shalom & Robert Cook-Deegan, *Patents and Innovation in Cancer Therapeutics: Lessons from CellPro* (2002) (unpublished study on file with authors).

122. See, e.g., Rai, *supra* note 33, at 115 (finding that “both communalism and norms against secrecy have been eroded by delays in publication and restrictions on the sharing of [biotech] research materials and tools caused by concerns about intellectual property rights,” but recognizing some reluctance to claim property rights in certain upstream discoveries by major research universities).

123. Rebecca Eisenberg, *Patenting Research Tools and the Law*, in NATIONAL RESEARCH COUNCIL, *IPR AND THE DISSEMINATION OF RESEARCH TOOLS IN MOLECULAR BIOLOGY 2* (1997) [hereinafter Eisenberg, *Patenting Research Tools*] (noting that, as a result of the Bayh-Dole Act, “institutions that perform fundamental research have an incentive to patent the sorts of early stage discoveries that in an earlier era would have been dedicated to the public domain”); see also STOKES, *supra* note 35, at 58-59 (regarding the lessening of well-defined distinctions between “basic” and “applied” research in certain areas).

124. Obstacles include potentially high direct and transactions costs between publicly funded institutions.

investigators. This prospect, in turn, puts institutional pressures on the sharing ethos in academia, which can limit disclosure and delay the publication of research results. It can also result in decisions not to publish at all or efforts to control the supporting data even after publication.

As databases become an ever more valuable commodity on the ledgers of university technology licensing offices, the university administrations may also become increasingly ambivalent about the established policies that promote open access to government-funded data.¹²⁵ Universities and segments of the academic scientific community tend to view proposals for new statutory protection of databases with considerable interest, although they also have expressed concerns about the deleterious effects that the wrong type of protection might have on education and research.¹²⁶ In addition, as a result of increased partnerships with private-sector firms, some universities have adopted stricter institutional rules and guidelines pertaining to access, use, and distribution of research data and data products.

2. The Zone of Informal Data Exchanges

The norms and policies described above characteristically apply to big science, where well-defined federal research programs and institutional structures implement the sharing ethos, and access to public domain data is facilitated by government-controlled repositories. However, small science presents a rather different picture, one in which individual investigators and small teams working independently predominate. As noted in Part I, here there are likely to be other, non-federal sources of funding and the federal support that is available will be less prescriptive about the terms of data availability, especially with respect to pre-publication data sets. Moreover, for research involving human subjects, strong regulations protecting personal privacy add an extra layer of secrecy.¹²⁷

In the experimental or laboratory sciences, such as chemistry or biomedical research, scientists use large databases for advancement to a much lesser extent than the observational sciences do, depending instead on the use of individual, repeatable experiments. The laboratory sciences also rely on the use of highly evaluated data sets and on published scientific literature, rather than on raw observational data. Because of the extremely specialized, labor-intensive nature of evaluated data sets, many are produced outside the government and are made available in proprietary publications or databases. Nevertheless, some public domain government sources exist for these types of data, even though they are smaller in number and volume than the sources of observa-

125. Cf. Eisenberg, *Bargaining*, *supra* note 13, at 235-39 (discussing universities' dilemma).

126. H.R. 354 Before the House Subcomm. on Courts and Intellectual Property, 106th Cong. (Mar. 18, 1999) (testimony of Charles Phelps on behalf of the Association of American Universities, the American Council on Education, and the National Association of State Universities of Land-Grant Colleges).

127. 5 U.S.C. § 552(a) (2000). The implementing regulations can be found at 45 C.F.R. § 56 (2002).

tional data.¹²⁸

The small science independent-investigator approach traditionally has also characterized a large area of fieldwork and studies, such as biodiversity, ecology, microbiology, soil science, and anthropology. Here, too, many individual or small-team data sets or samples are independently collected and analyzed.¹²⁹ The data from such studies generally have been heterogeneous and unstandardized, with few of the individual data holdings deposited in public data repositories or openly shared.¹³⁰

a. Normative Conflicts

In the small science environment, individual investigators working autonomously are more driven by self-interest and a competitive spirit. They are one step removed from the top-down policies that regulate big science projects because in most cases they are not dependent on access to and participation in large public domain data repositories or other structured research programs that formally promote and enforce the sharing ethos. In this realm, scientists have much greater freedom to limit the disclosure of their data, in keeping with their private interests, at least until their research is formally published. Even then—because few institutionalized public data repositories exist—they may choose to withhold all but the minimum amount of data needed to support their published findings, or may provide no data at all.¹³¹

128. See BITS OF POWER, *supra* note 1, at 52. Examples of evaluated laboratory physical sciences data include the Evaluated Nuclear Structure Data File, and various materials science and chemical sciences data. *Id.* at 205-12.

129. See generally NATIONAL RESEARCH COUNCIL, FINDING THE FOREST IN THE TREES: THE CHALLENGE OF COMBINING DIVERSE ENVIRONMENTAL DATA (1995) [hereinafter FINDING THE FOREST].

130. *Id.*; see also BITS OF POWER, *supra* note 1, at 83-88. However, as discussed *infra* in the last section of Part II, the advent of pervasive distributed computing and digital networks has led to the organization of many areas of previously “small science” into “big science” types of initiatives. Notable examples include the Human Genome Project and several ecological and biodiversity programs with networked and partially centralized data resources. The U.S. Long Term Ecological Research Network (“LTER Net”) now connects more than 1100 scientists and students investigating ecological processes at twenty-four research sites. Data are freely available within two to three years. See LTER Net, <http://lternet.edu> (last visited Jan. 10, 2003). The Global Biodiversity Information Facility (“GBIF”), the purpose of which is to “make the world’s biodiversity data freely and universally available” through an interoperable network of biodiversity databases and information technology tools, provides another example. See GBIF, <http://www.gbif.org/> (last visited Jan. 10, 2003).

131. In a recent national survey, “forty-seven percent of geneticists who asked other faculty for additional information, data, or materials regarding published research reported that at least [one] of their requests has been denied in the preceding [three] years. Ten percent of all post publication requests for additional information were denied. Because they were denied access to data, [twenty-eight percent] of geneticists reported that they had been unable to confirm published research. Twelve percent said that in the previous [three] years, they had denied another academician’s request for data concerning published results.” Eric G. Campbell, et al., *Data Withholding in Academic Genetics: Evidence from a National Survey*, 287 JAMA 473-80 (2002). For a similar situation in neuroscience research, see Peter Aldhous, *Prospect of Data Sharing Gives Brain Mappers a Headache*, 406 NATURE 445, 445-46 (2000), describing how proposals to make data sharing a mandatory requirement for publication produced a significant negative response from some members of the research community. See generally Jon Cohen, *Share and Share Alike Isn’t Always the Rule in Science*, SCIENCE, June 21, 1995, at 1715.

This lack of structure, however, may paradoxically make single scientists more dependent on cooperative relationships with other scientists working in the same or related disciplines to expand their data supplies and integrate data that result from disparate investigations. To this end, they are prompted to voluntarily construct informal sharing arrangements with other scientists, notwithstanding the rivalry and competitive goals that divide them and despite the economic opportunities that can make such disclosures risky.¹³² These arrangements give rise to collective streams of data that flow through informal networks of voluntary collaborations built from the bottom up.¹³³

Data in this sphere have a different configuration from that which one is accustomed to seeing in the more formal, big science endeavors. There, the data resources are often centralized databases from which scientists will borrow bits for their research, which they may combine with other similar data resources, or with personally collected data, into individual data products. In the informal, mostly small science zone of data exchanges, however, data resemble a continuous stream or assemblage that has to be constructed from the bottom up and that flows in different directions and at different rates as informal networks evolve. These data streams consist of “chains of products” in which “increasingly refined materials and inscriptions downstream” play an important role.¹³⁴

Individual investigators constantly take parts of data sets from others and alter or refine them in ways that suit their own needs. These value-added outputs are then exchanged with the value-adding contributions of other investigators. As Stephen Hilgartner and Sherry Brandt-Rauf describe it:

Any end-point may become a starting point; every output may become an input. One therefore cannot assume that data somehow arrive . . . on the scene in pre-packaged units that are transferable, shareable, or publishable, or that there is some discrete point in time at which data should naturally be transferred. On the contrary, there is always more than one way of dividing a data stream into portions that may or may not be disseminated.¹³⁵

In this process timing is crucial because many of these entities are novel or “extremely scarce” and become available only by special arrangements, if at all.

¹³⁶

132. Cf. Richard Nelson, *The Market Economy, and the Republic of Science* 17 (July 24, 2000) (unpublished draft, on file with authors) (“[T]he fact that most of scientific knowledge is open, and available through open channels, is extremely important. This enables there to be at any time a significant number of individuals and firms who possess and can use the scientific knowledge they need in order to compete intelligently in this evolutionary process. The ‘communalism’ of scientific knowledge is an important factor contributing to its productivity in downstream efforts to advance technology.”).

133. See Stephen Hilgartner & Sherry I. Brandt-Rauf, *Data Access, Ownership, and Control: Toward Empirical Studies of Access Practices*, 15 *KNOWLEDGE: CREATION, DIFFUSION, UTILIZATION* 355, 355-72 (1994) [hereinafter Hilgartner & Brandt-Rauf, *Access, Ownership & Control*] (describing the “data stream” dynamics in biomedical research); see also Hilgartner, *Access to Data*, *supra* note 31; Hilgartner & Brandt-Rauf, *Controlling Data*, *supra* note 31.

134. Hilgartner & Brandt-Rauf, *Access, Ownership & Control*, *supra* note 133.

135. *Id.*

136. *Id.*

Of course, individuals and small groups operating in this zone may nucleate.¹³⁷ For example, “[e]nvironmental information systems frequently nucleate around informal collaborations (including volunteers) that determine useful partnerships.”¹³⁸ If these collaborations continue to evolve on a sharing basis, they may give rise to more formal government research program structures that will then institutionalize the data-sharing protocols on an open access foundation.¹³⁹

However, the sharing that occurs to produce the data streams in many areas of small science, particularly in the context of the laboratory life sciences, arises from a delicate process of barter and exchange. Here access to data is both limited and targeted, a process that is mediated by intricate inter-, intra-, and extra-laboratory negotiations.¹⁴⁰ In this process, autonomy and dependence foster conflicting interests. By withholding data, single scientists or laboratories temporarily promote their individual comparative advantages. At the same time, they may lose opportunities to improve their positions by gaining mediated access to others’ data.¹⁴¹ In making decisions and negotiating barter or exchange transactions, the scientist’s trading mentality is tempered to some extent by the sharing ethos, especially as expressed in peer pressure, but he or she may be impervious at this stage, at least, to the formal institutional mechanisms that might otherwise enforce that ethos.

Accordingly, it is important to reiterate that the contents of the data streams we describe here fall largely outside the public domain that was previously described, even though the bulk of the data that function as inputs into the process have been funded by the government. Rather, because these operations are proceeding on an informal, non-organized basis (typically in a pre-publication environment), the data are likely to be held under varying degrees

137. See Powell, *Networks of Learning*, *supra* note 13, at 265 (“In fields such as biotech, where knowledge is advancing rapidly and the sources of knowledge are widely dispersed, organizations enter into an array of relationships to gain access to different competencies and knowledge.”).

138. RESOLVING CONFLICTS, *supra* note 8, at 73.

139. Indeed, this is exactly what appears to be occurring in the areas of ecological studies and biodiversity, which, until recently, were conducted by individuals or small groups in autonomous field studies in separately funded programs. The data collected in these investigations were heterogeneous, unstandardized, lacking in rigorous data management protocols, and generally not shared or made available for many years, if at all. FINDING THE FOREST, *supra* note 129, at 84-96, 100-01. With the advent of the Internet, however, many of these previously disparate and autonomous research groups have begun to share their data through formally organized networks with formal, standardized protocols. See, e.g., the organizations discussed at *supra* note 130.

140. Hilgartner & Brandt-Rauf, *Access, Ownership & Control*, *supra* note 133, at 369. In contrast to open publication, Hilgartner and Brandt-Rauf suggest that access here is achieved by a variety of means, including barter; selected distribution to colleagues; patents; training; confidential sharing; purchase and sale transactions; “pre-release” to corporate sponsors; or the data is held in lab for future uses.

141. See Stephen Hilgartner, *Data Access Policy in Genome Research*, in PRIVATE SCIENCE 202-15 (Arnold Thakray ed., 1998) (giving examples of data-sharing within the Human Genome Project).

of actual secrecy, much like know-how in classical industrial property relations.¹⁴²

The problem is that for the process of mutually beneficial exchange to work, individual investigators need access to others' published and unpublished data sets, and all the players in the relevant community find themselves in a similar position. While the pressure to cooperate is obvious and is reinforced by the norms of science, self-interest—including economic self-interest and the competitive spirit—may pull strongly in the opposite direction. The system as a whole would clearly benefit in the long term from fruitful cooperative arrangements. However, there is no guarantee that individual scientists will be willing to absorb the short-term losses to which disclosure subjects them, nor is it certain that these losses will in fact be offset by future gains from cooperation.¹⁴³ In sum, there are strong pressures to hold out or, at least, to hold back and to divulge no more than is necessary in the process of barter and exchange.

How well the process works in practice depends on a number of factors. One factor is the communications networks themselves, which may or may not provide low-cost and efficient access to the different distributed sources of data. In the pre-digital environment, it was much harder to build these collaborative networks owing to constraints of space and time, whereas the Internet now collapses these limits and provides instantaneous, low-cost opportunities to overcome previous barriers to data sharing in small-scale research.¹⁴⁴ A second factor is the extent to which the relevant community or sub-community respects the sharing ethos and imposes peer pressure to enforce it. A third factor is the degree to which commercial pressures from private sector partners and universities themselves distort the sharing ethos and stimulate profit-maximizing strategic behavior in academics.¹⁴⁵ A fourth factor is the extent to which direct personal opportunities to profit either from the economic value of the data sets, their reputational value, or a combination of both will drive each player's decisions and further undermine the cooperative ethos. Of course, many additional idiosyncratic factors that may affect personal behavior and choices will come into play as well.

There is much about this small science process that is still unknown outside the circles of discrete research collaborations or subdisciplines because its dis-

142. Cf. STEVEN P. LADAS, PATENTS, TRADEMARKS, AND RELATED RIGHTS 1616-74 (1975) ("The International Protection of Know-How").

143. See Eisenberg, *Patenting Research Tools*, *supra* note 123, at 7 ("Negotiating for access to research tools might present particularly difficult problems for would-be licensees who do not want to disclose the direction of their research in its early stages by requesting licenses.").

144. See discussion of the LTER Net and GBIF examples, *supra* note 130.

145. See DAVID, NATURE'S SECRETS, *supra* note 8, at 10 ("In their dual capacities the administrators of academic institutions (and the individuals who staff them) must continue to seek effective ways of mediating conflicts between the societal goals that will be served by preserving the organizational modes and norms of open scientific inquiry, on the one hand, and, on the other hand, the lure of capturing for their more immediate and private purposes a larger portion of the 'information rents'—by circumscribing free access to the new knowledge gained through the researches conducted under their auspices.").

tinctive sociology has only recently become the subject of serious scholarship, and because different scientific communities operate with different value structures and have diverse needs. In addition, one should not overstate the relative importance of this informal zone in relation to the more formal structures described above, which are particularly dominant in basic research.¹⁴⁶

What can be said is that much of science depends on this zone of informal data exchanges, it processes more and more data of great economic value to industry, and the value of its data streams can be greatly increased through the use of digital networks and more refined legal tools. At the same time, an increasingly protectionist legal regime and the corresponding pressures to commercialize make the voluntary nature of these exchanges ever more precarious and subject to hoarding and holdout tendencies. An already fragile, informal process of barter and exchange that largely depends on a balance of personal self-interest and the cooperative norms of science is coming under intense economic and legal pressures that call its continued vitality into question.

b. A Different Legal Regime

A salient characteristic of the zone of informal data exchanges is the extent to which it has operated without reliance on formal legal rules. Intellectual property rights, including copyrights, have played no supporting role, although patent law—as discussed below—has increasingly cast an encroaching shadow over data transactions in which it previously played a minor part. Contractual agreements like those that regulate relations between funding entities and investigators in the formal zone have seldom been used in the informal zone. Such rules that have existed seem to derive largely from unfair competition law, and the participating scientists appear to have been much less aware of them than of the relevant norms of science, as expressed through peer pressure.

In the informal zone, scientists collaborate on a voluntary, cooperative basis to produce what amounts to an inchoate common data resource, but their access to it depends on the individual transactions they are able to negotiate. Certain uses of the data and data products that comprise this resource may be regulated by the granting agency's rules and the norms of science. Yet, until the research results are published, the investigators typically will have enjoyed a period of exclusive use conferred by the grants, and the grantees are not otherwise subject to supplementary formal rules deriving either from intellectual property rights or contractual arrangements with peers. Even after publication, the federal grants usually exhort the grantee to make the underlying data available, without imposing hard and fast obligations to do so. The granting agencies have, in any case, found it difficult to enforce even these moral obligations so long as the data remain outside a public repository and under the control of

146. These relationships exist primarily at the interface of science and technology, or where the line of demarcation between basic and applied research has collapsed. Their importance varies by type of investigation.

the principal investigator. Many other sources of research funding impose no such obligations at all. The upshot is that the scientist in the informal zone retains considerable legal discretion in determining the amount and even the conditions of disclosure, subject of course to peer pressure and his or her own need to barter for access to the cumulative data stream.

In this raw state of affairs, the cumulative collection, although potentially of much greater value than the sum of its parts, remains vulnerable to strategic manipulation in keeping with the opportunities that single scientists have to commercialize or otherwise personally exploit some or all of it. Viewed individually, these scientists will normally have made no formal commitment to promote the collective interest in an inchoate data commons, and all members of the community are wary of the opportunities for strategic behavior.

Moreover, none of the players really knows what the others actually possess, so that the totality of the emerging common resource cannot be self-consciously designed (as in, say, Linux),¹⁴⁷ but is overwhelmingly dependent on single decisions to collaborate and reveal—decisions that are costly and could be zero sum.¹⁴⁸ The vulnerability of the cumulative output to strategic behavior impedes the extent to which any given player will be willing wholeheartedly to commit his data holdings to the enterprise. Thus, risk aversion is high, and there are few if any formal legal tools to reduce it.

This is not to say that law is totally absent from this environment. To be more accurate, it is present in the form of legal rules that protect know-how and confidential information generally, as a subcategory of industrial property law. At bottom, these rules put limits on the worst forms of strategic behavior, such as espionage and deception, in case the norms of science otherwise fail to impede such activities.¹⁴⁹ They also provide a source of quasi-moral rights, very important to scientists, by impeding a scientist from passing off another's data contribution as his or her own.¹⁵⁰

Moreover, the loose liability rules¹⁵¹ that are embodied in trade secret and unfair competition law give some incentive to cooperate in informal exchanges

147. See discussion *infra* at Part IV.C.1.b.

148. See generally Hilgartner, *Access to Data*, *supra* note 31 (discussing zero-sum competition situations in the context of scientific practice).

149. RESTATEMENT (THIRD) OF UNFAIR COMPETITION §§ 39-45 (1995) [hereinafter RESTATEMENT UNFAIR COMPETITION].

150. Because the raw values in factual data compilations remain non-copyrightable in the United States, such information is in the public domain, free for the taking. *Feist Publ'ns, Inc. v. Rural Tel. Servs. Co.*, 499 U.S. 340 (1991). While the source should be acknowledged as a matter of professional ethics and etiquette, there is no legal recourse for the data originator against an unacknowledged use by another scientist, except as unfair competition law allows. Such uses probably occur quite frequently, and in many cases the use is either undetectable or the originator of the data does not mind. However, unacknowledged uses sometimes erupt in a very public way. See Eliot Marshall, *DNA Sequencer Protests Being Scooped With His Own Data*, 295 SCIENCE 1206, 1206-07 (2002). In foreign countries, the moral rights of copyright law could be invoked. See Berne Convention, *supra* note 109, at art. 6 *bis*.

151. "Whenever someone may destroy the initial entitlement if he is willing to pay an objectively determined value for it, an entitlement is protected by a liability rule." Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1092 (1972). Under so-called property rules (in this case, the exclusive rights of intellectual

because these rules cease to protect data that are independently discovered or revealed without improper appropriation from another. Unfair competition laws do not confer an exclusive property right, but rather a set of liability rules that impede certain market-destructive ways of depriving any individual player's data set of its value by improper means.¹⁵² But these liability rules will not impede any other player from independently creating the same data by proper means, and if the need is great enough, this may occur. When it does, any exchange value that the original data holder may have enjoyed will be diminished or eliminated. In this sense, actual or legal secrecy confers a kind of property right—a right against improper behavior and perhaps a certain amount of lead time—but it is also a “disappearing right” that becomes vulnerable to discovery and disclosure by others.¹⁵³

Even this incentive to cooperate, however, may be much reduced in science, as compared to technical innovation, for the reason that it may not be possible for a second scientist to independently recreate the data set in question, or it may be prohibitively expensive and inefficient to do so.¹⁵⁴ In such cases, the generally weak liability rules governing confidential information can produce too much protection for data held in actual secrecy because there is, in practice, virtually no functional equivalent of reverse engineering.¹⁵⁵ When that happens, the temptation to hoard and holdout can be high, and the corresponding costs of cooperation are similarly elevated.

While the legal weight of these liability rules should not be underestimated, they are clearly of limited value in the zone of informal data exchanges. Probably, their greatest relevance will be at the moment when commercial opportunities are most palpable and when wholesale appropriation might produce tangi-

property law), one cannot take the entitlement in question without prior permission of the owner. In this sense, property rules are “absolute permission rules.” Robert P. Merges, *Institutions for Intellectual Property Transactions: The Case of Patent Pools*, in EXPANDING THE BOUNDARIES OF IP, *supra* note 13, at 123, 131 (2001) [hereinafter Merges, *Patent Pools*]. “By contrast, liability rules are best described as ‘take now, pay later.’ They allow for non-owners to take the entitlement without permission of the owner, so long as they adequately compensate the owner later.” *Id.*

152. See *Bonito Boats, Inc. v. Thunder Craft Boats, Inc.*, 489 U.S. 141, 167-68 (1989) (invalidating Florida “plug mold” statute to protect against copying of boat hull designs); *Compco Corp. v. Day-Brite Lighting, Inc.*, 376 U.S. 234, 237-38 (1964) (state unfair competition laws not to protect unpatented lamp designs in absence of source confusion); *Sears, Roebuck & Co. v. Stiffel Co.*, 376 U.S. 225, 231-32 (1964) (preventing use of Illinois unfair competition law to block copying of an unpatented lamp design). Whether these rules impede “copying” or wholesale appropriation as such remains an open question. Compare *Int'l News Serv. v. Assoc. Press*, 248 U.S. 215 (1918) and *Nat'l Basketball Ass'n v. Motorola, Inc.*, 105 F.3d 841 (2nd Cir. 1996) with *Bonito Boats*, 489 U.S. at 141 and *Wal-Mart Stores, Inc. v. Samara Bros.*, 529 U.S. 205 (2000). For the view that wholesale copying should be prohibited, see, for example, Wendy J. Gordon, *On Owning Information: Intellectual Property and the Restitutory Impulse*, 78 VA. L. REV. 149, 165 (1992), proposing a tort of “malcompetitive copying,” and Dennis S. Karjala, *Misappropriation as a Third Intellectual Property Paradigm*, 94 COLUM. L. REV. 2594, 2601-08 (1994) [hereinafter Karjala, *Misappropriation*].

153. John C. Stedman, *Trade Secrets*, 23 OHIO ST. L.J. 4, 21 (1962) (characterizing trade secret rights as “disappearing rights”).

154. See A QUESTION OF BALANCE, *supra* note 1, at 29-30 (discussing the uniqueness of many scientific and technical databases).

155. See generally Pamela Samuelson & Suzanne Scotchmer, *The Law and Economics of Reverse Engineering*, 111 YALE L.J. 1575 (2002).

ble economic gains. It is precisely here, however, where the traditional rules governing confidential information may be the least potent, because they cannot usually be triggered without a showing of legal, not merely actual, secrecy, and the tenets of legal secrecy will often be too high for academic scientists to meet.¹⁵⁶ Absent some general norm against the wholesale appropriation of another's data production,¹⁵⁷ the ability of scientists cooperating in the zone of informal data exchanges to invoke rules based on unfair competition law to preserve their individual and collective patrimonies remains speculative at best.

The most salient feature of the legal infrastructure as it impinges on this zone is, therefore, the extent to which it has been largely irrelevant to the social and scientific processes underway. In other words, if the available legal mechanisms do little to support the fragile process of barter and exchange, they also create few barriers to that process. However, this situation could change radically with the introduction of new technological means for controlling digital data, new opportunities for commercial exploitation, and above all, the enactment of new intellectual property rights in non-copyrightable collections of data as discussed in Part III.

C. Private-Sector Data as a Component of the Research Commons

The private sector is both a user of public domain data and a major producer of them. The contribution of the private sector to the research commons under traditional legal rules has been under-appreciated. There is also a tendency within the academic community to overlook the needs of the private sector for access to public domain data and, accordingly, a tendency to underestimate the effects that restrictions on access to such data may have on private sector research and development ("R&D").

The private sector generates an ever-increasing amount of scientific data that are indispensable to academic research. Yet, more and more scientific data produced by academics come freighted with restrictions and conditions that arise from partnerships between academic institutions and industry. How to organize the research commons so that both academic and private firms can obtain the data they need without compromising either their public or private pursuits is a key issue for science policy and this article.¹⁵⁸

1. Dissemination and Use of Scientific Data in Private Sector R&D

The private sector is a major producer of scientific data that enters the public domain under existing legal rules. Disregarding database publishers for

156. See, e.g., David Friedman et al., *Some Economics of Trade Secret Law*, 5 J. ECON. PERSP. 61, 63-66 (1991). Of course, "passing off" another's data set as one's own would not require legal secrecy. See RESTATEMENT UNFAIR COMPETITION, *supra* note 149.

157. State unfair competition laws may or may not provide such a norm, depending on how courts interpret *International News Service v. Associated Press*, 248 U.S. 215 (1918). See Hilgartner & Brandt-Rauf, *Access, Ownership & Control*, *supra* note 133.

158. See Part IV, *infra*, for an extensive analysis of this dilemma and the authors' proposed approaches for addressing it.

a moment, private sector researchers frequently publish their findings, like academics, in which case the traditional rules of copyright law apply as previously described, and there is a corresponding duty to release at least enough non-copyrightable data as are needed to support the published results. If the research project were funded by the federal government, there would typically be further obligations to disclose or deposit the underlying data sets following publication or completion of the project. If the research leads to a patent application, then that application must also meet minimum utility and disclosure requirements.¹⁵⁹ If the research results are neither published nor patented and are held under actual or legal secrecy, then any resulting innovations may be reverse engineered by honest means.¹⁶⁰ In that case, the technical information or “know-how” derived from the process of reverse engineering should also enter the public domain.¹⁶¹

Besides generating scientific data as a by-product of their R&D activities, the private sector may affirmatively contribute data directly to the public domain for a number of different reasons. For example, companies may donate data sets accumulated over time that are of value to science, even if they no longer possess significant direct commercial value to the firm. Companies may also self-publish scientific and technical information to promote their goods and services. They increasingly may decide to put even commercially valuable technical data in the public domain to deter competitors from blocking fruitful lines of research by strategic patent applications derived from the data thus disclosed.¹⁶²

Private sector R&D activities, of course, make use of both government-generated data and government-funded data, especially from academic research undertaken at universities. Of greater importance, however, is the vast information commons that consists of the cumulative and sequential know-how and the state-of-the-art, which the community at work on any given technical trajectory has acquired and routinely shares over time.¹⁶³ This know-how—that is, information about how to achieve commercial advantages by sub-patentable

159. 35 U.S.C. §§ 101, 102, 111, 112 (2000) (utility, novelty, disclosure, and enablement requirements in patent law).

160. UNIF. TRADE SECRETS ACT (USTA) § 1(4), 14 U.L.A. 449 (1985).

161. For a recent discussion, see Samuelson & Scotchmer, *supra* note 155. See also J. H. Reichman, *Overlapping Proprietary Rights in University-Generated Research Products: The Case of Computer Programs*, 17 COLUM.-VLA J.L. & ARTS 51, 93-98 (1992).

162. See, e.g., SNP Consortium, available at <http://snp.cshl.org> (last visited Feb. 14, 2003), discussed *infra* notes 676-78; see also Douglas Lichtman et al., *Strategic Disclosure in the Patent System*, 53 VAND. L. REV. 2175, 2197-99, 2204-09 (2000); Rai, *supra* note 33, at 112-13 (discussing reluctance of leading research universities to patent expressed sequence tags (“ESTs”) and single nucleotide polymorphisms (“SNPs”). For an explanation of SNPs, see *infra* note 676.

163. See, e.g., Margaret Sharp, *Technological Trajectories and Corporate Strategies in the Diffusion of Biotechnology*, in TECHNOLOGY AND INNOVATION: CRUCIAL ISSUES FOR THE 1990S 93, 94-97 (Erico Deiacco, et al, eds., 1990) 221-35 (Giovanni Dosi et al. eds., 1988); Richard R. Nelson, *Intellectual Property Protection for Cumulative Systems Technology*, 94 COLUM. L. REV. 2674, 2676 (1994) (distinguishing “traditional discrete invention model” from “cumulative systems model”).

technical means—is typically acquired by trial and error and shaped by investors in R&D into incremental innovation products.¹⁶⁴

The information commons that underlies and pervades the realm of sub-patentable innovation is a fundamental asset of a competitive economy based on constant innovation.¹⁶⁵ In principle, patents are not granted unless the would-be inventor exceeds the level of innovation that the community at work on a given technical trajectory would be expected to produce. Below this line of “non-obviousness,”¹⁶⁶ trade secret laws do not impede employees from carrying their skills and personal know-how—their personal quotient of the information commons—from one job to another. The spillover effects this produces, and the rapid interchange of information between members of the technical community it encourages, become the engine of small-scale innovation in Silicon Valley and its equivalents around the world.¹⁶⁷

In recent years, however, the free flow of information between members of the engineering community at work on given technical trajectories has increasingly been slowed and disrupted by a proliferation of hybrid intellectual property rights that fall between the patent and copyright paradigms.¹⁶⁸ Beginning with industrial design laws and utility model laws, which date back to the nineteenth century, examples include plant breeders’ rights, integrated circuit designs, and most recently, *sui generis* rights in non-copyrightable databases.¹⁶⁹

164. J. H. Reichman, *Of Green Tulips and Legal Kudzu: Repackaging Rights in Subpatentable Innovation*, 53 VAND. L. REV. 1743, 1747-53 (2000) [hereinafter Reichman, *Of Green Tulips*]; see also Frischmann, *supra* note 70. This information is neither copyrightable nor patentable, but is usually kept under actual secrecy. If, instead, investors spend the money to keep their valuable know-how under legal secrecy, unfair competition law will protect it against misappropriation by improper means—such as industrial espionage—but not against reverse-engineering by honest means. The second-comer’s duty to reverse-engineer the processes by which innovative sub-patentable products are made gives the investor a period of natural lead time in which to recoup investments and establish a trademark. See UTSA, *supra* note 160, at § 1(4); J. H. Reichman, *Legal Hybrids Between the Patent and Copyright Paradigms*, 94 COLUM. L. REV. 2432, 2511-20 (1994) [hereinafter Reichman, *Legal Hybrids*] (discussing chronic shortage of natural lead time under present-day conditions).

165. Cf. James Boyle, *Cruel, Mean or Lavish? Economic Analysis, Price Discrimination and Digital Intellectual Property*, 53 VAND. L. REV. 2007 (2000) (stressing benefits of competition and need for public domain inputs into the information economy) [hereinafter Boyle, *Cruel, Mean or Lavish*].

166. 35 U.S.C. § 103 (2000).

167. See Samuelson & Scotchmer, *supra* note 155; see also Rochelle Cooper Dreyfus, *Trade Secrets: How Well Should We Be Allowed to Hide Them? The Economic Espionage Act of 1996*, 9 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1 (1998) (criticizing the federal criminal Trade Secrets Act for adverse impact on spillover effects).

168. Reichman, *Legal Hybrids*, *supra* note 164.

169. In many countries, utility model laws protect functional designs and small-scale innovations generally. See, e.g., UMA SUTHERSANEN, *DESIGN LAW IN EUROPE* 383-93 (2000) (discussing various European utility model laws); Mark D. Janis, *Second Tier Patent Protection*, 40 HARV. INT’L L.J. 151, 155-59 (1999) (classifying and discussing classical utility model regimes). While the United States does not have a utility model law, Congress has opted to protect two sets of functional designs. See Vessel Hull Design Protection Act of 1998, 17 U.S.C. §§ 1301-32 (2000); Semiconductor Chip Protection Act of 1984, 17 U.S.C. §§ 901-914 (2000). Most countries, other than the United States, have enacted *sui generis* laws to protect non-functional industrial designs. See, e.g., SUTHERSANEN, *supra*, at 28-54; Graeme B. Dinwoodie, *Federalized Functionalism: The Future of Design Protection in the European Union*, 24 AM. INTELL. PROP. L. ASS’N Q.J. 611 (1999); J. H. Reichman, *Design Protection in Domestic and Foreign Copyright Law: From the Berne Revision of 1948 to the Copyright Act of 1976*, 1983 DUKE

These regimes are enacted because those who invest in applications of know-how to industrial products—"incremental innovation bearing know-how on its face"—become increasingly vulnerable to second-comers who duplicate their products without expending the time and money to reverse engineer or independently create. In supplying investors with artificial lead time, however, these hybrid regimes tend to block follow-on applications and disrupt the information commons that had driven sub-patentable innovation in the past.¹⁷⁰ As discussed in detail in Part III, the greatest threat to this information commons and the innovation systems it supports is the new hybrid exclusive property right in non-copyrightable databases, which has been promulgated throughout Europe by the 1996 Directive on the legal protection of databases.¹⁷¹ This Directive provides a new and potentially perpetual intellectual property right in exchange for mere investment, which is qualitatively different and more far-reaching in its impact on the public domain than any traditional intellectual property regime.¹⁷²

For present purposes, it suffices to note that because relations between the private sector and universities have flourished under the influence of the Bayh-Dole Act and related legislation,¹⁷³ the private sector on the whole tends to make greater use of data generated by academic research than in the past. As these two sources of data become increasingly commingled, there is a growing conflict between the open access norms of traditional university research and the tendency of the private sector to restrict access to and use of data by means of legal secrecy and the aforementioned hybrid exclusive property rights.¹⁷⁴ In this connection, the European database right will not only affect the distribution of academic data itself, but will also enable the private sector to restrict access to and use of all the data within its control in ways that were not previously possible. Any efforts to organize the scientific information commons on a more rational basis must accordingly take account of the role that the private sector plays both as generator and user of scientific and technical data in research.

2. Database Publishers

One would not need to treat database publishers in a separate entry so long as they remained subject to the same legal regime applicable to academic scientists described above. Under that regime, traditional copyright rules applied

L. J. 1143 (1983). The United States, instead, continues to rely on design patents, 35 U.S.C. § 171 (2000). For discussion of *sui generis* database protection laws, see *infra* Part III.B.2.d.

170. See generally Reichman, *Of Green Tulips*, *supra* note 164, at 149-56; J. H. Reichman, *Computer Programs as Applied Scientific Know-How: Implications of Copyright Protection for Commercialized University Research*, 42 VAND. L. REV. 639, 656-69 (1989).

171. See *supra* note 17.

172. See *infra* text accompanying notes 347-73.

173. See, e.g., Rai, *supra* note 33, at 110 (discussing growth of academic-industrial relationships in the past two decades).

174. Cf. Rai, *supra* note 33, at 110-11 ("Participants in these academic-industrial relationships often depart quite markedly from traditional research norms.").

only to the original and creative selections and arrangements of compilations of data.¹⁷⁵ As a result, the bulk of the data entered the public domain unless otherwise protected either by state trade secret laws, or to a still unknown extent, state unfair competition laws prohibiting the wholesale duplication of non-copyrightable databases.¹⁷⁶

It is worth reiterating that the U.S. database industry has thrived in this environment in which weak protection has been the rule, and it has commanded a large share of the world database market.¹⁷⁷ To sustain this success, the database industry relies on the constant updating of its databases and new releases, which add value as well as business protection to previously accumulated data, and on a pronounced “niche market” effect, which makes it difficult for second-comers to enter a given market segment once the first-mover has established a solid reputation for quality. In the networked environment, investors also rely heavily on technological fences to protect their databases; electronic, standard-form license contracts for mass markets; and negotiated licenses with institutional customers to regulate access to and use of their databases.¹⁷⁸

While the contractual regulation of online dissemination of databases poses serious problems for science and education overall, the underlying dimensions of the public domain are shrinking rapidly owing to new and unprecedented legislative initiatives, such as the E.C. Database Directive¹⁷⁹ and similar database protection proposals in the United States.¹⁸⁰ To glimpse the far-reaching repercussions of this type of legislation, one has only to consider its potential impact on the notion of publication as the line of demarcation between private and public ownership of scientific and technical information. Under traditional copyright law and practices, a scientist who published an article will have dedicated any supporting data that accompanied it to the public domain. If the *sui generis* database right applied, however, academics could publish their articles, and they or their publisher could still retain the legal rights to control or restrict use of the same data even after publication.¹⁸¹

How the academic community would ultimately react and adjust to this new legal situation would depend on many different factors,¹⁸² and some of them are considered later in this article. The point for now is that the established portrait of the public domain for published collections of data is in a state of flux. The radical changes stemming from legal and technical measures tending to fence that domain make it advisable for the scientific community to consider the need to construct and manage a research commons of its own for the dissemination

175. See *supra* notes 85-91 and accompanying text.

176. See *supra* note 150 and accompanying text.

177. See, e.g., Maurer, *Across Two Worlds*, *supra* note 34.

178. See *supra* note 34.

179. See *supra* note 17.

180. See *supra* notes 17, 26 and accompanying text.

181. In practice, scientific use would then depend to some extent on explicit exceptions or immunities built into the relevant database protection laws.

182. See generally Reichman & Uhler, *Database Protection*, *supra* note 4.

of scientific and technical data and information that heretofore automatically entered the public domain.

D. Potentially Enhanced Role of Public Domain Data in a Digitally Networked Research Environment

Because science builds upon science, the production of data sets is not an end in itself, but rather the means to an end, the first step in the creation of new information, knowledge, and understanding. As part of that process, the original databases are continually refined and recombined to create new databases and new insights. Each level of processing adds value to an original or raw set of data by summarizing its contents, providing different interpretations of their meaning, or synthesizing new information products. As Nobel laureate Joshua Lederberg testified before a congressional committee considering the enactment of a new database protection bill:

Data are the building blocks of knowledge and the seeds of discovery. They challenge us to develop new concepts, theories, and models to make sense of the patterns we see in them. They provide the quantitative basis for testing and confirming theories and for translating new discoveries into useful applications for the benefit of society. They are also the foundation of sensible public policy in our democracy. The assembled record of scientific data and resulting information is both a history of events in the natural world and a record of human accomplishment.¹⁸³

The primary purpose of this section is to emphasize the extent to which digital technologies have revolutionized the role that data in the public domain or available under open access policies can play in the research process. To quote Lederberg once again:

[The] recent advent of digital technologies for collecting, processing, storing, and transmitting data has led to an exponential increase in the size and number of databases created and used. A hallmark trait of modern research is to obtain and use dozens or even hundreds of databases, extracting and merging portions of each to create new databases and new sources for knowledge and innovation.¹⁸⁴

As will become readily apparent, the successful implementation of these data integration functions depends to a large extent on the availability, access to, and unrestricted use of affordable data resources in the public domain.

1. Digitally Pooled Data Resources

The enhanced role that digital technologies play in government-sponsored data collection activities¹⁸⁵ constantly adds new dimensions to both big and small science. Increasingly powerful sensor technologies are unmasking layers of previously hidden attributes of our natural universe and documenting their essential characteristics in automated data streams. The resulting data sets

183. *Collections of Information Antipiracy Act: Hearing on H.R. 354 Before the House Comm. on the Judiciary*, 106th Cong. 189-205 (March 18, 1999) (statement of Joshua Lederberg, President, Rockefeller University, on behalf of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine, and American Association for the Advancement of Science) [hereinafter *Collections Hearings*].

184. *Id.*

185. *See supra* Part II.A.1 (outlining these activities).

become “foundational” in the sense that they establish a baseline characterization of natural objects or processes that then becomes a common resource for research in that particular area. Other data sets track natural phenomena or behavior on a longitudinal basis over long time periods.

In the observational sciences, for example, space-based and ground-based sensors continue to collect vast amounts of digital data about our planet and outer space in all regions of the electromagnetic spectrum.¹⁸⁶ Moreover, the miniaturization of sensor technologies is making it possible to place thousands of tiny sensor arrays in different ecosystems to collect environmental data concerning physical, chemical, and biological processes.¹⁸⁷ This instrumenting of the environment using both remote and embedded network sensing devices makes the pervasive monitoring of the ecosphere possible, with the resulting data sets able to be archived in public data centers or made directly available in the digitally networked environment for a broad range of long-term studies and applications.

Similarly, the aforementioned large facilities in the experimental physical sciences, such as nuclear fusion, high-energy laser, and neutron-beam devices, create ever-larger amounts of data that are processed and analyzed for new discoveries and applications.¹⁸⁸ The data sets generated by each of these large observational and experimental facilities and related data centers have grown from gigabyte levels just a decade ago, to terabytes and, in some cases, petabytes of data currently.¹⁸⁹

Formally structured big science programs also increasingly rely on the compilation of large, public domain databases composed of the contributions of hundreds of individual investigators from government, academia, and even industry, who generate data independently and then contribute their findings to government or government-funded data centers. Many of these types of arrangements are in the rapidly growing area of bioinformatics, such as molecular biology,¹⁹⁰ ecological,¹⁹¹ and biodiversity studies,¹⁹² as well as in numerous other areas of research. Both the centralized and distributed public domain data repositories constitute a common resource from which all researchers can borrow freely, whether for fundamental exploratory investigations or for more specifically applied problem-solving purposes.

The continuing advancement of observational sensors and experimental data production technologies, particularly in miniaturized desktop computa-

186. See *supra* note 45 and accompanying text.

187. See Gregory Bonito, *Emergent Sensor Technologies*, in *SCALABLE INFORMATION NETWORKS FOR THE ENVIRONMENT* (Alison Withey et al. eds., 2002).

188. See *supra* note 45 and accompanying text.

189. The data center at the high-energy physics research center, CERN, and at the U.S. Geological Survey's Earth Resources Observing Systems Data Center, which archives various land remote sensing data, are now approaching the petabyte level (petabyte = quadrillion bytes).

190. See, e.g., National Center for Biotechnology Information, *supra* note 45.

191. See discussion of LTER Net, *supra* note 130.

192. See discussion of GBIF, *supra* note 130.

tional devices or portable instruments for use in fieldwork, also increasingly empowers single scientists to collect or create their own data sets in the informal, small science domain. As digital capabilities for autonomous data production improve and proliferate, an immense new supply of potentially interconnected information is being created within each discipline community, which complements the large, formally structured, public-domain foundational data sets emanating from the big science programs. Although the availability of individually created scientific information resources situated in the informal research domain remain subject to the cultural and sociological impediments described above,¹⁹³ these resources nonetheless constitute a rapidly growing corpus of potentially available inputs into the research commons.

Turning from collection to dissemination functions, the growing ability of each community to use the Internet to provide virtually universal access to all this information—assuming it remains otherwise available—has revolutionized the conduct of scientific research. U.S. researchers in all scientific and engineering areas are among the most active users of the Internet. This is hardly surprising given the primary role of the U.S. governmental and academic science and technical community in the original development and early evolution of the Internet.¹⁹⁴ What may be less apparent, however, is that the architecture and function of the Internet itself arose as a technological manifestation of many of the same cultural attributes that characterize the public research enterprise. The Internet was developed by publicly funded government and academic researchers as a voluntary and cooperative integration of highly distributed autonomous networks, operating in a largely self-regulated and international system, through which information could pass freely and without restrictions.¹⁹⁵

These parallel attributes of the public research community and the early Internet provided a natural impetus to scientists to use digital networks immediately and pervasively to make access to their reciprocal inputs and outputs generally available.¹⁹⁶ All of the formally organized public scientific data repositories now make their holdings known and available through the Internet, although direct access to the full set of their archived information tends to be limited by technical or security factors.¹⁹⁷

193. *See supra* Part II.B.2.

194. *See, e.g.*, Vinton Cerf, *How the Internet Came to Be*, in *THE ONLINE USER'S ENCYCLOPEDIA* (Bernard Adoba ed., 1993).

195. *Id.*; *see also* LAWRENCE LESSIG, *CODE AND OTHER LAWS OF CYBERSPACE* (1999) (discussing Internet architecture and its initial open design).

196. In the informal sphere, however, this reciprocity is often the product of negotiated exchanges.

197. The principal technical factor that limits direct access to large databases is insufficient bandwidth in many current Internet connections, although this is expected to change rapidly with the introduction of "grid" technology, particularly in the research community. Security-based technical protection of online databases against malicious hackers further limits direct access to the entire content and creates other inefficiencies. *See generally* NATIONAL RESEARCH COUNCIL, *TRUST IN CYBERSPACE* (Fred B. Schneider ed., 1999).

The opportunities for direct peer-to-peer exchanges of data, and for new distributed research collaborations on either a formal or informal basis may be an even more important development.¹⁹⁸ Particularly noteworthy in the context of this article is the stimulus and means that the Internet provides for increasing the flow of the inchoate data stream that characterizes the informal zone of data exchanges. By providing a direct, instantaneous, and relatively secure means to communicate and share data, digital networks potentiate unlimited opportunities for implementing the cooperative and sharing ethos that has been fundamental to the progress of science. When most of one's professional peers within and outside a given area of specialization openly post their findings on their websites or remain willing to exchange information on request, the potential for serendipitous results and synergistic advances becomes greatly increased.¹⁹⁹

Moreover, these integrating benefits of the Internet are not limited to single peer-to-peer exchanges, important as they may be. Ubiquitous digital networks also make possible entirely new forms of organized peer production, from small groups working in distributed laboratories²⁰⁰ to network-wide, volunteer-based, open modes of production that may have particularly significant implications and applicability in the public science context.²⁰¹ Certainly, voluntary, collaborative peer production of data, information, and knowledge was a hallmark trait of science long before the digital era. The co-authoring of research articles, joint compilations of databases, sharing of knowledge and research results at public conferences, peer-review process for publications, and participation in large research programs over long time periods were all "peer production" activities. These norms and practices of the pre-digital era take on a heightened importance in the digitally networked environment and lead to the possibility for greatly expanded peer production opportunities ideally suited to open, public research.²⁰²

2. Electronic Transformation Tools

While digital data production technologies and networks allow scientists to pool and instantaneously exchange the raw materials of their research, advanced software and ever more powerful computational tools enable them to process, organize, and transform the raw data into discoveries and applications. The following highlights three notable advances in this context.

A key development is the use of new software tools to process and integrate large amounts of data to create refined data products and applications. Software enables scientists to take portions of pre-existing databases from different

198. See Yochai Benkler, *Coase's Penguin, or, Linux and the Nature of the Firm*, 112 YALE L.J. 369 (2002) [hereinafter Benkler, *Coase's Penguin*].

199. See discussion of network effects, *infra* notes 223-27 and accompanying text.

200. See generally NATIONAL RESEARCH COUNCIL, COLLABORATORIES: IMPROVING RESEARCH CAPABILITIES IN CHEMICAL AND BIOMEDICAL SCIENCES (1999).

201. See Benkler, *Coase's Penguin*, *supra* note 198; see also DAVID, GOOD FENCES, *supra* note 20, at 3-4.

202. See Benkler, *Coase's Penguin*, *supra* note 198.

sources in order to combine and reprocess them to address complex problems. A quintessential example is geographic information systems (“GIS”) software, which provides the means of integrating diverse environmental and socio-economic data on geospatially referenced grids for a broad array of basic research goals and practical applications.²⁰³

Many of the data sources used in this process are governmental or government-funded at the federal, state, and local levels, including some of the key foundational data sets mentioned above. The U.S. Geological Survey estimates that geospatial data applications in the United States alone contribute over \$3.5 trillion annually to the economy.²⁰⁴ The successful utilization of GIS depends in large part on easy access to the relevant data at affordable prices and, especially, on access with few restrictions on reuse and redissemination. Conversely, data sources that are too expensive or come freighted with many user limitations can undermine or even block a research project or important application. Of course, the same issues arise in other areas of data intensive research, such as biotechnology, econometrics, or climate change studies.

Data mining techniques, also known as knowledge discovery in databases, provide the means to extract salient data from very large databases and automatically convert the extracted components into useful information or even new discoveries.²⁰⁵ For example, data mining algorithms were used to discover twenty new quasars in a huge astronomical database in just a few hours of processing time. A search of the same database by a team of astronomers would have taken forty times longer to yield the same results.²⁰⁶ Performing such a search and extracting those discoveries could constitute an infringement under the E.C. Database Directive.

The final technology highlighted here is grid computing, which may be defined as a “super Internet” for high-performance computing by integrating geographically and organizationally dispersed computational resources, such as CPUs, storage systems, communication systems, data sources and instruments, and the researchers themselves.²⁰⁷ By providing “pervasive, dependable, consistent, and inexpensive” access to such advanced computational capabilities and resources, researchers believe that computational grids will have a transforming effect similar to the electric power grid a century ago, which would allow new

203. For a basic description of GIS functions, see U.S. Geological Survey, at <http://www.usgs.gov/research/gis/title.html> (last visited Feb. 14, 2003). See also Environmental Systems Research Institute, Inc., at www.gis.com/whatisgis/index.html (last visited Jan. 10, 2003).

204. FEDERAL GEOGRAPHIC DATA COMMITTEE, REPORT OF THE CIVIL IMAGERY AND REMOTE SENSING TASK FORCE ON THE VALUE OF CIVIL IMAGERY AND REMOTE SENSING 2 (October 1, 2002).

205. Definition adapted from *Introduction to Data Mining*, available at http://www.andypryke.com/university/dm_docs/dm_intro.html (last visited Jan. 10, 2003) (listing a variety of background resources on this technology).

206. See Usama Fayyad, *Industrial Keynote Address: Data Mining and Databases*, in DATA FOR SCIENCE AND SOCIETY: THE SECOND NATIONAL CONFERENCE ON SCIENTIFIC AND TECHNICAL DATA (2000), available at http://books.nap.edu/html/codata_2nd/ch15.html (last visited Feb. 14, 2003).

207. Definition adapted from *What is a Grid?*, available at <http://www.aei.mpg.de/~manuela/Gridweb/info/grid.html> (last visited Jan. 10, 2003).

classes of applications to emerge.²⁰⁸ An example of an initial project using grid technology is the European Union's "Data Grid." This initiative is expected to "enable next generation scientific exploration which requires intensive computation and analysis of shared large-scale databases, millions of Gigabytes, across widely distributed scientific communities."²⁰⁹ The selected applications areas include high-energy physics, biomedical research, and satellite earth observations.²¹⁰

Taken together, the new digital technologies discussed above, as well as numerous others omitted for reasons of brevity, enable scientists to perform the following quantitatively and qualitatively new functions:

- Collect and create unprecedented and ever increasing amounts and types of raw data about all natural objects and phenomena;
- Collapse the space and time in which data and information can be made available;
- Facilitate entirely new forms of distributed research collaboration and information production; and
- Interpret and transform the raw data into unlimited new configurations of information and knowledge.

However, the successful implementation of all these functions remains heavily dependent on the continued existence of a robust public domain and related open access policies and practices to realize the promise of new research and innovation tools.

III

SCIENTIFIC DATA AS A PRIVATE GOOD: PRESSURES ON THE PUBLIC DOMAIN AND THEIR IMPLICATIONS

The map described in the preceding pages suggests the extent to which the vast reservoir of accumulated public domain data feeds into the scientific research infrastructure and the system of innovation to which it gives rise. It has often been pointed out that government-supported research plays a primary role in making this system of innovation the most productive in the world. However, the indispensable role of public domain data in nourishing this system, while generally taken for granted, is less clearly understood or appreciated, and new possibilities for further potentiating this role in the digital environment constitute another "endless frontier." These matters require more attention lest growing pressures to fence the scientific commons lead to unforeseen and unintended consequences that could seriously disrupt the national system of innovation as a whole.

208. THE GRID: BLUEPRINT FOR A NEW COMPUTING INFRASTRUCTURE xvii (Ian Foster & Carl Kesselman eds., 1999). See also Ian Foster's web site, at <http://www-fp.mcs.anl.gov/~foster> (last visited Feb. 14, 2003) (listing numerous Grid technology resource documents).

209. E.U. Data Grid Project, at <http://web.datagrid.cnr.it> (last visited Jan. 10, 2003).

210. Learn More about DataGrid, <http://web.datagrid.cnr.it/LearnMore/index.jsp> (last visited Jan. 10, 2003).

A. Shifting the Public–Private Boundaries

The prominent role that public domain data have played in fueling American science and innovation is partly explained by the economic literature concerning public goods. Public goods, unlike private goods, are characterized by their *nonrival* and *nonexcludable* properties.²¹¹ The former means that it costs nothing to provide the good to another person once someone has produced it, that is, it tends to have zero marginal cost. The latter means that once such a good has been produced, the producer cannot exclude others from benefiting from it. Typical examples of public goods that fully satisfy both criteria are national defense and the operation of lighthouses.

The problem that public goods pose is that however important their production may be for the political body as a whole, they will attract insufficient or suboptimal private investment because investors cannot deter free-riding second-comers or otherwise fully recoup the return from their investments.²¹² Governments typically respond to this problem by making the investments that the private sector cannot or should not be expected to provide, such as funding public health and safety objectives, or the national defense.²¹³ Often these public initiatives will generate substantial additional private investment and supplementary social benefits as the private sector finds ways to convert upstream public investments in public goods into downstream commercial products and services that do generate appropriable returns.²¹⁴

From this perspective science itself, especially basic science, resembles a public good, which private enterprise could not adequately support.²¹⁵ For example, research on tropical diseases, climate change, and astrophysics falls into the category of global public goods, as do large fundamental research facilities, such as space science spacecraft missions and high-energy particle accelerators for physics experiments, for which high up-front costs and noncommercial or uncertain applications make public investment the only feasible alternative. These public investments, in turn, contribute to the “knowledge infrastructure” required for efficient R&D directed at exploitable commercial innovations.²¹⁶

211. Inge Kaul et al., *Defining Global Public Goods*, in GLOBAL PUBLIC GOODS: INTERNATIONAL COOPERATION IN THE 21ST CENTURY (Kaul et al. eds., 1999).

212. See generally ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 108-18 (3d ed. 2000).

213. RESOLVING CONFLICTS, *supra* note 8, at 23 (“Most government functions are carried out by the public sector either because of an overriding public interest in the outcome, or because the potential for high risk or low payoff makes the task unattractive to the private sector.”).

214. See, e.g., Frischmann, *supra* note 70, at 357-60 (stressing tension between maximizing consumption of public goods and constraining consumption to maximize market-based efficiency).

215. Michael Callon, *Is Science a Public Good?*, 19 SCI. TECH. & HUM. VALUES 395, 400 (1994) (“The qualification of science as a quasi-public good rather than as a full-fledged public good derives essentially from the fact that it is to a certain degree appropriable—whereas in standard theory a true public good has to be completely inappropriable.”).

216. Paul David, *The Political Economy of Public Science*, in THE REGULATION OF SCIENCE AND TECHNOLOGY 38 (Helen Lawton Smith ed., 2001) [hereinafter David, *Political Economy*]. According to a recent study, some seventy-three percent of all patents granted in the United States during the

Notable examples include Internet communications protocols, the global positioning system, and computer simulation methods for the visualization of molecular structures.²¹⁷

As Paul David and Michael Callon have reminded us,²¹⁸ there are good reasons for protecting many scientific endeavors from competitive market forces that cannot efficiently allocate resources for the production and distribution of pure public goods.²¹⁹ Because industry and business tend to under-invest in scientific production, government takes up the slack either by intervening directly or by providing incentives to the private sector to overcome market failure, in the form of legal monopolies falling within the domestic and international intellectual property systems.

To appreciate the full implications of the map drawn in Part II, moreover, it is important to establish that scientific and technical data also manifest the properties of “quasi public goods” that economists associate with science as a whole. Information, facts, and ideas, once divulged, cost nothing to propagate and become difficult to keep from others. Thus, the government intervenes to promote potential long-term economic and social benefits because the data produced from basic scientific research are often too commercially risky to be developed by the private sector.²²⁰

Even when scientific data emerge from applied science, where partial appropriability and foreseeable returns attract private investment, they may present public good properties that justify government support. For example, governments invest in providing advanced weather data largely because of the need to ensure public safety and the protection of the nation’s economic assets. However, by electing to provide such data without intellectual property protection, either free or at no more than the marginal cost of dissemination, the U.S. government goes beyond its mission of protecting the safety of its citizens and their property to providing the raw material for a dynamic value-adding private sector. As scholars have noted:

As a result of this concept of public/private partnership, the U.S. boasts a robust private meteorology industry with revenues in excess of \$500 million annually, and a rapidly growing weather risk management industry with risk management instruments approaching a value of \$8 billion. The authors believe that the relatively small size of

1990s cited government or government-funded research. Francis Narin et al., *The Increasing Linkage Between U.S. Technology and Public Science*, 26 RES. POL’Y 317 (1997).

217. “Much more critical over the long run than ‘spin-offs’ from basic science programmes are their cumulative indirect effects in raising the rate of return on private investment proprietary R&D performed by business firms.” David, *Political Economy*, *supra* note 216, at 39.

218. *Id.* at 35; Callon, *supra* note 215.

219. David, *Political Economy*, *supra* note 216, at 36 (“[T]he findings of scientific research, being new knowledge, would be seriously undervalued were they sold directly through perfectly competitive markets.”).

220. “Under U.S. policy, most federal government data are in the public domain and cannot be copyrighted. By making data easy and inexpensive to obtain, the U.S. government seeks to promote science, create a more informed public, and foster the development of a thriving commercial information industry.” RESOLVING CONFLICTS, *supra* note 8, at 24.

these sectors in the E.U. is primarily due to the restrictive data policies of a number of governments and their national meteorological services.²²¹

Appropriate public sector investment in the production of scientific data benefits from what economists call “positive externalities” and “network effects.” A positive externality occurs when one party confers benefits on another without the latter having to fully compensate the former.²²² Basic research, together with the creation and dissemination of scientific databases, especially in their raw form, may have no immediate economic applications or market, but they can lead subsequently to unanticipated or serendipitous advances and whole new spheres of innovation and commerce. Such activities provide prime examples of positive externalities that direct government support can greatly promote and that may not be undertaken at all without such support.

The other important concept here is that of a network effect, which arises when the value of using a particular type of product depends on the number of users.²²³ Examples of products with high positive feedback from network effects include telephones and fax machines, if there are many users rather than only a few.²²⁴ Perhaps the quintessential product with positive network effects is the Internet. From this perspective, scientific databases or other collections of information can add considerably more value to society and the economy if they are openly available on the Internet (assuming that production remains feasible in the absence of appropriability, as would occur with government funding).²²⁵

Scientists were the pioneers of the Internet revolution and have become some of the most prolific users of the medium for accessing, disseminating, and using data.²²⁶ When data are provided as a public good via the Internet, unencumbered by proprietary rights, the positive externalities from network effects can be especially high. They become even greater to the extent that the data are prepared and presented in a way that makes them available and usable to a broader range of non-expert users outside the scientific community.

As Joseph Stiglitz and his colleagues point out:

The shift toward an economy in which information is central rather than peripheral may thus have fundamental implications for the appropriate role of government. In

221. Plujimers & Weiss, *supra* note 47, at 7 (citing authorities).

222. An externality may be defined as the action of one entity affecting the well-being of another, without appropriate compensation. A negative externality is the imposition of additional costs by entity A (for example, through the deleterious effects of pollution created by A) on entity B, without A's having to pay for those costs. Conversely, a positive externality confers benefits (innovation) from A to B without full compensation to A. JOSEPH E. STIGLITZ ET AL., COMPUTER & COMMITTEE INDUSTRIAL ASSOCIATION, THE ROLE OF GOVERNMENT IN A DIGITAL AGE 33 (2000).

223. *Id.* at 42.

224. S. J. Liebowitz & Stephen E. Margolis, *Network Externality: An Uncommon Tragedy*, J. ECON. PERSP., Spring 1994, at 133, 133-36 (giving further examples of network effects).

225. See Benkler, *Coase's Penguin*, *supra* note 198; cf. Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CAL. L. REV. 479 (regarding the question of adequate incentives in peer production projects).

226. See *supra* notes 194-95 and accompanying text.

particular, the public good nature of production, along with the presence of network externalities and winner-take-all markets, may remove the automatic preference for private rather than public production. In addition, the high fixed costs and low marginal costs of producing information and the impact of network externalities are both associated with significant dangers of limited competition.²²⁷

These economic characteristics associated with the transmission of digital scientific data on the Internet provide a strong argument for many of the activities previously described, which are undertaken within the public domain by government agencies or by non-governmental entities receiving government support.

At the same time, attention to the economic literature suggests the importance of ascertaining the limits of public good analysis lest government compete with or otherwise undermine activities that the private sector could carry out more efficiently.²²⁸ Scientific data, like much of science itself, are not a pure public good to the extent that they can be bundled and embodied in physical artifacts that make them appropriable to a certain degree.²²⁹ Information can come in two forms: codified knowledge and incorporated knowledge. The former may be expressed in a standardized and compact form, so as to permit easy, low-cost transmission, verification, storage, and reproduction. The latter, by contrast, is inscribed in some machine.²³⁰

From this perspective, scientific data and information are both inputs into the national system of innovation and outputs of that system, and intellectual property rights tend to regulate the flow of private sector investments into downstream applications of research data.²³¹ How this balance of public and private interests worked out in the past was largely reflected in the map of public domain data depicted in Part II. Whatever imperfections that map brings to light, its most salient feature is that the balance of interests it reflects underlies the world's most productive and successful system of innovation.²³²

However, that inherently dynamic and shifting balance of interests has come under intense pressure in recent years for a number of different reasons. The "convergence technologies" that greatly improve access to information also afford "technological means of inhibiting access in ways that were never before practical."²³³ Global competition has induced governments in developed countries to strengthen existing intellectual property rights, enact new and more

227. STIGLITZ ET AL., *supra* note 222, at 44.

228. See *supra* Part II.A.2. (discussing the limits on competition with the private sector by the federal government).

229. See, e.g., Callon, *supra* note 215, at 398 (stressing that pure public goods are completely inappropriable).

230. *Id.* at 397.

231. See David C. Mowery & Nathan Rosenberg, *The U.S. National System of Innovation*, in NATIONAL SYSTEMS OF INNOVATION—A COMPARATIVE ANALYSIS 29-75 (Richard R. Nelson ed. 1993); Reichman & Franklin, *supra* note 16, at 884-86 (discussing dual function of information in the networked environment).

232. See Mowery & Rosenberg, *supra* note 231, at 47-54, 59-64.

233. DAVID, DIGITAL BOOMERANG, *supra* note 20, at 10; see also Reichman & Franklin, *supra* note 16, at 897-99 (restored power of the "two-party" deal in digital environment).

powerful rights not previously experimented with, and to press for high levels of harmonized protection at the international level.²³⁴ Of particular importance in the United States are efforts by government “to cut expenditures by transferring to the private sector a range of data production and information distribution activities that formerly were publicly provided.”²³⁵ The end result has been the collapse of the established lines of demarcation between public and private interests that were codified in the classical patent and copyright paradigms,²³⁶ and the enclosure and transformation of “larger and larger portions of the public data ‘commons’ . . . into private monopolies.”²³⁷

The question this raises is the extent to which the functions of public domain data on which science and innovation have traditionally relied, as illustrated above, may be compromised by ill-conceived initiatives to stimulate investment for short-term private gain without sufficient attention to the long-term needs of both science and innovation, as well as the broader society. To this end, the remainder of this Part of the article will summarize the economic and legal assaults on the research commons that have recently occurred, and examine some of the implications of those pressures on the ability of the commons to continue to perform both its traditional and potentially enhanced functions in the digitally networked environment.

B. Pressures on the Research Commons

The digital revolution has made investors acutely aware of the heightened value that collections of data and information may acquire in the new information economy.²³⁸ Attention has logically focused on the incentive and protective structures for generating and disseminating digital information products, especially online. Although most of the legal and economic initiatives have been focused on—and driven by—the entertainment sector, software producers, and large publishing concerns, there is growing interest in the possibility that commoditization of even public sector and public domain data could stimulate substantial investments by providing new means of recovering the costs of production.²³⁹ Moreover, investors have increasingly understood the economic

234. See, e.g., KEITH E. MASKUS, INTELLECTUAL PROPERTY RIGHTS IN THE GLOBAL ECONOMY 1-14, 15-85 (2000); Peter Drahos, *Developing Countries and International Intellectual Property Standard-Setting*, 5 J. WORLD INTELL. PROP. 765, 769-83 (2002).

235. DAVID, DIGITAL BOOMERANG, *supra* note 20, at 8.

236. J. H. Reichman, *Charting the Collapse of the Patent-Copyright Dichotomy*, 13 CARDOZO ARTS & ENT. L.J. 475 (1993).

237. DAVID, DIGITAL BOOMERANG, *supra* note 20, at 11.

238. See, e.g., Maurer, *Across Two Worlds*, *supra* note 34.

239. See, e.g., *The Collections Of Information Antipiracy Act and the Vessel Hull Design Protection Act : Hearing on H.R. 2652 Before the Subcommittee on Courts and Intellectual Property of the House Comm. on the Judiciary 105th Cong. (1997)* (testimony by Laura d'Andrea Tyson) [hereinafter *Information Antipiracy Hearings*]. This testimony was based on a research project funded by Reed-Elsevier, Inc., and The Thomson Corp., completed Sept. 5, 1997. See also G. M. Hunsucker, *The European Database Directive: Regional Stepping Stone to an International Model*, 7 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 697 (1997); YALE M. BRAUNSTEIN, ECONOMIC IMPACTS OF DATABASE PROTECTION IN DEVELOPING COUNTRIES AND COUNTRIES IN TRANSITION (W.I.P.O. Standing

potential that awaits those who capture and market data and information as raw materials or inputs into the upstream stages of the innovation process.²⁴⁰

What follows focuses first on pressures to commoditize data in the public sector and then on legal and technological measures that endow database producers with new proprietary rights and novel means of exploiting the facts and data that copyright law had traditionally left in the public domain. These pressures arise both within the research community itself and from forces extraneous to it. How that community responds to these pressures over time will determine the future metes and bounds of the information commons that supports scientific endeavors.

If, as we have reason to fear, current trends will greatly diminish the amount of data available from the public domain, this decrease could initially compromise the scientific community's ability to fully exploit the promise of the digital revolution. Moreover, if these pressures continue unabated and become institutionalized at the international level,²⁴¹ they could disrupt the flow of upstream data to both basic and applied science and undermine the ability of academia and the private sector to convert cumulative data streams into innovative products and services.

The pressures discussed below also pose serious conflicts between the norms of public science and the norms of private industry. We contend that failure to resolve these conflicts and properly balance the interests at stake in preserving an effective information commons could eventually undermine the national system of innovation.²⁴²

1. Commoditization of Data in Public Science

During the last ten years, there has been a marked tendency to shift the production of science-relevant databases from the public to the private sector. This development occurred against the background of a broader trend in which the government's share of overall funding for research and development vis-à-vis that of the private sector has decreased from a high of sixty-seven percent in the 1960s to twenty-six percent in 2000.²⁴³ Furthermore, since the passage of the Bayh-Dole Act in 1980, the results of federally funded research at universities have increasingly been commercialized either by public-private partnerships

Committee on Copyright and Related Rights, #SCCR/7/2, 2002), available at http://www.wipo.int/eng/meetings/2002/sccr/pdf/sccr7_2.pdf (last visited Jan. 10, 2003).

240. There were, of course, always concerns about incentives to produce basic data and information as raw materials of the innovation process, especially in light of perceived gaps in intellectual property law that seemed to leave databases in limbo. See, e.g., Denicola, *supra* note 78.

241. See J. H. Reichman, *Database Protection in a Global Economy*, *supra* note 18, 485-500 (managing transnational database protection without harmonization).

242. Cf. Mowery & Rosenberg, *supra* note 231, at 47-51, 53-56, 62-64 (describing the interplay of public-private interests in the U.S. system of innovation); Richard R. Nelson & Nathan Rosenberg, *Technical Innovation and National Systems*, in NATIONAL INNOVATION SYSTEMS, *supra* note 231, at 3, 5-9 (stressing the extent to which science and technology are intertwined).

243. Industry and Agency Concerns over Intellectual Property Rights, Testimony before the Subcomm. on Technology and Procurement Policy, House Comm. on Government Reform, 107th Cong. (May 10, 2002) (testimony of Jack L. Brock, Jr., United States General Accounting Office).

with industry or directly by the universities themselves.²⁴⁴ Industry support of university research has increased in certain sectors, such as medical research, even as the federally funded share of university research support has declined.²⁴⁵

a. Reducing the Scope of Government-Generated Data

The budgetary pressures on the government are both structural and political in nature. On the whole, mandated entitlements in the federal budget, such as Medicare and Medicaid, are politically impossible to reduce, and as their costs mount, the money available for other discretionary programs, including federally sponsored research, has shrunk as a percentage of total expenditures.

This structural limitation is compounded by the rapidly rising costs of state-of-the-art research, including some researcher salaries, scientific equipment, and major facilities. With specific regard to the information infrastructure, researchers typically earmark the lion's share of expenses to computing and communications equipment, with the remainder devoted to managing, preserving, and disseminating the public domain data and information that results from basic research and other federal data collection activities. The government's scientific and technical data and information services are thus the last to be funded and are almost always the first to suffer cutbacks.

For example, the National Oceanic and Atmospheric Administration's ("NOAA") budget for its National Data Centers remained flat and actually decreased in real dollars between 1980 and 1994, while its data holdings increased exponentially and the overall agency budget doubled (mostly to pay for new environmental satellites and a ground-based weather radar system that are producing the exponential data increases).²⁴⁶ Information managers at most other science agencies have complained about reductions in funding for both their data management and scientific and technical information budgets.²⁴⁷

These chronic budgetary shortfalls for managing and disseminating public domain scientific data and information have been accompanied by recurring

244. See, e.g., Nelson, *supra* note 132; Rai, *supra* note 33, at 110-11; see also Mowery & Rosenberg, *supra* note 231, at 53 (stressing that closer ties between industry and universities restored a linkage that had been weakened in the 1950s and 1960s).

245. The percentage of university medical research funded by the federal government decreased from approximately seventy-five percent in 1976 to approximately sixty-four percent in 1997. Between 1992 and 1999, the percent of industry funding in the same research sector increased from just under seven percent to just under eight percent. Hamilton Moses III, *Academic Relationships with Industry, Presentation to the Government, University, Industry Research Roundtable* (Mar. 28, 2001) (on file with authors).

246. See *BITS OF POWER*, *supra* note 1, at 62. This trend has continued for NOAA up to the present time. However, in 1996, Congress authorized NOAA to charge fair market value for its data and to institute a two-tiered pricing system with discounts for educational organizations that place their orders for data online. See *supra* note 56. Since then, revenues from data sales have decreased about six percent per year. NATIONAL OCEANIC & ATMOSPHERIC ADMINISTRATION, U.S. DEPARTMENT OF COMMERCE, *THE NATION'S ENVIRONMENTAL DATA: TREASURES AT RISK* 35-36 (2001).

247. Interviews with STI managers in the U.S. Geological Survey, the Department of Agriculture, the Department of Energy, and the Department of Defense (2002).

political pressures on the scientific agencies to privatize their outputs.²⁴⁸ Until recently, for example, the common practice of the environmental and space science agencies was to procure data collection systems, such as observational satellites or ground-based sensor systems, from private companies. Such procurements were typically made under cost-plus contracts and pursuant to government specifications based on consensus scientific requirements recommended by the research community.²⁴⁹ Private contractors would build and deliver the data collection systems, which the agencies would then operate pursuant to their mission. All data from the system would belong to the government and would enter the public domain.

Today, however, industry has successfully pursued a strategy of providing an independent supply of the government's needs for data and information products rather than building and delivering data collection systems for government agencies to operate.²⁵⁰ This solution leaves the control and ownership of the resulting data in the hands of the company and allows it to license them to the government and to anyone else willing to pay. Because of this new-found role of the government agency as cash cow, there has recently been a great deal of pressure on the science agencies, particularly from Congress, to stop collecting or disseminating data in-house and to obtain them from the private sector instead.

This approach previously resulted in at least one well-documented fiasco, namely, the privatization of the NASA–NOAA Landsat earth remote sensing program in 1985, which seriously undermined basic and applied research in environmental remote sensing in the United States for the better part of a decade.²⁵¹ More recently, the Commercial Space Act of 1998 directed the National Aeronautics and Space Administration (“NASA”) to purchase space and earth science data collection and dissemination services from the private sector and to treat data as commercial commodities under federal procurement regulations.²⁵² The meteorological data value-adding industry has directed similar lobbying pressures at NOAA.²⁵³ The photogrammetric industry has likewise

248. In other countries, notably those in the European Union, it has been a longstanding policy and practice to commercialize data right from the public source. *See generally* Green Paper, *supra* note 46; Plujimers & Weiss, *supra* note 47.

249. The various discipline boards and committees of the National Academies are requested by the federal science agencies periodically to provide research strategies for specific discipline or research program areas. *See, e.g.*, NATIONAL RESEARCH COUNCIL, ASTRONOMY AND ASTROPHYSICS IN THE NEW MILLENNIUM (2001); NATIONAL RESEARCH COUNCIL, RESEARCH STRATEGIES FOR THE U.S. GLOBAL CHANGE RESEARCH PROGRAM (1990), *available at* <http://www.nap.edu> (last visited Jan. 10 2003).

250. *See, e.g.*, Management Association for Private Photogrammetric Surveyors (“MAPPS”), Licensing Data, Licensing People, summary proceedings of November 2000 conference, *available at* <http://www.mapps.org/library.asp> (last visited Jan. 10, 2003) [hereinafter MAPPS Conference].

251. For a discussion of the adverse effects that the privatization of the Landsat program had on basic research, see BITS OF POWER, *supra* note 1, at 121-24.

252. Commercial Space Act of 1998, Pub. L. No. 105-303, §§ 107(a), 107(b), 112 Stat. 2843, 2853 (1998).

253. *See, e.g.*, *Hearing Before the House Science Comm., Subcomm. on Energy and the Environment*, 105th Cong. (1997) (statement of Michael S. Leavitt, President, Weather Services Corp., on behalf of

indicated a desire to expand the licensing of data products to the U.S. Geological Survey and to other federal agencies.²⁵⁴

Efforts have also been made by various industry groups to limit the online information dissemination services of several federal science and technology agencies. In the cases of the patent database of the U.S. Patent and Trademark Office, the PubMed Central database of peer-reviewed life science journal literature (provided on a free and unrestricted basis by the NIH National Library of Medicine), and certain types of weather information disseminated by the National Weather Service, such efforts have proved unsuccessful to date. However, publisher groups did succeed in terminating the Department of Energy's *PubScience* web portal for physical science information.²⁵⁵

b. Commercial Exploitation of Academic Research

Turning to government-funded research activities, the trend of greatest concern for purposes of this article is the progressive incorporation of data and data products into the commercialization process that is already underway in academia. The original purpose of the Bayh-Dole Act and related legislation was primarily to enable universities to obtain patents on applications of research results.²⁵⁶ More recently, this activity has expanded to securing both patents and copyrights in computer programs. Now, databases used in molecular biology have themselves become sources of patentable inventions, and the potential commercial value of these databases as research tools has attracted considerable attention and controversy.²⁵⁷

These and other databases are increasingly the subject of licensing agreements prepared by university technology transfer offices, which may be prone to treat databases like other objects of material transfer agreements.²⁵⁸ The

the Commercial Weather Services Association); *Contracting Out and Privatization Opportunities in NOAA: Hearing Before the Senate Governmental Affairs Comm., Subcomm. on Oversight Government Management and the Dist. of Columbia*, 105th Cong. (1997) (statement of Joel Myers, President, AccuWeather, Inc.).

254. MAPPS Conference, *supra* note 250.

255. *Goodbye, PubScience, We Hardly Knew Ye: Free DOE Database Goes Dark*, LIBR. J. ACAD. NEWSWIRE, Nov. 12, 2002. It should also be noted that this lobbying effort was supported vigorously in 2001 by the American Chemical Society, a major scientific society publisher.

256. Eisenberg, *Public Research*, *supra* note 120, at 1665.

257. *See, e.g.,* DAVID & FORAY, *supra* note 3, at 16-17 (“[C]ooperatively assembled bioinformatic databases are permitting researchers to make important discoveries in the course of ‘unplanned journeys through information space.’ If that space becomes filled by a thicket of property rights, then those voyages of discovery will become more expensive to undertake . . . and the rate of expansion of the knowledge base is likely to slow.”); Powell, *supra* note 13, at 254-55, 263-65, (“But what is striking is how actively universities and firms are seeking to privatize new information” in biotechnology and related fields.).

258. Empirical research has shown a growing acceptance by universities of confidentiality, nondisclosure, and other restraints on open research as a result of increasing private-sector partnerships. *See* W.M. Cohen et al., *Industry and the Academy: Uneasy Partners in the Cause of Technological Advance*, in CHALLENGES TO RESEARCH UNIVERSITIES (R. Noll ed., 1998). University technology transfer offices operate on a commercial business model and view other universities as competitors. *See* Jerry G. Thursby & Marie C. Thursby, *Who is Selling the Ivory Tower? Sources of Growth in University Licensing*, 48 MGMT. SCI. 90, 90 (2002) (indicating that there has been a dramatic increase in technol-

default rules that such licensing agreements tend to favor are exclusive arrangements under onerous terms and conditions that include restrictions on use, and even grant-back and reach-through clauses claiming interests in future applications.²⁵⁹

Moreover, there is a growing awareness in academic circles generally that data and data products may be of considerable commercial value, and individual researchers have become correspondingly more wary of making them as available as before.²⁶⁰ This trend, together with the pressures on government agencies described above, could pose serious problems for the research community's ability to access and use needed data resources under any circumstances.²⁶¹ In reality, these problems could become much greater as the new legal and technological fencing measures discussed below become more broadly implemented.

2. Intellectual Property, E-Contracts, and Technological Fences

Part II of this article showed that traditional copyright law was friendly to science, education, and innovation by dint of its refusal to protect either facts or ideas as eligible subject matter; by limiting the scope of protection for compilations and other factual works to the stylistic expression of facts and ideas; by carving out express exceptions and immunities for teaching, research, and libraries; and by recognizing a catch-all, fall-back fair use exception for non-profit research and other endeavors that advanced the public interest in the diffusion of facts and ideas at relatively little expense to authors. Reinforcing these policies were judge-made and partially codified exceptions for functionally dictated components of literary works, which take the form of non-protectible methods, principles, processes, and discoveries.²⁶² On the whole, these principles tended to render facts and data as such ineligible for copyright protection and allow researchers to access and use facts and data otherwise embodied in protectible works of authorship without undue legal impediments.

In contrast, recent legal developments in intellectual property and contracts law have radically changed the pre-existing regime. These and other related developments now make it possible to assert and enforce proprietary claims to virtually all the factual matter that previously entered the public domain the moment it was disclosed.

ogy transfer through licensing by universities as they attempt to appropriate returns from faculty research); *cf.* Rai, *supra* note 33, at 113-15 (noting failure of post-1995 efforts to develop uniform rules on biological materials transfer agreements to govern universities and private companies, and tendency of universities to depart significantly in practice from 1995 inter-university agreement).

259. See Stephen M. Maurer, Promoting and Disseminating Knowledge: The Public/Private Interface, paper prepared for the National Research Council's Symposium on the Role of Scientific and Technical Data and Information in the Public Domain 39-41 (Sept. 5-6, 2002), available at http://www7.nationalacademies.org/biso/Maurer_background_paper.html (last visited Jan. 10, 2002) (noting that approximately half of all university licensing agreements are on an exclusive basis) [hereinafter Maurer, Promoting and Disseminating Knowledge].

260. See, e.g., Campbell et al., *supra* note 131. See discussion *supra* Part II.B.1.c.

261. See, e.g., Powell, *supra* note 13, at 264-66.

262. See *supra* notes 11-13 and accompanying text; Part II.B.1.b.

Some of the earliest changes were intended to bring U.S. copyright law into line with long-standing norms of protection recognized in the Berne Convention. For example, the principle of automatic copyright protection, the abolition of technical forfeiture due to lack of formal prerequisites, such as notice, and the provision of a basic term of protection lasting for the life of the creator plus fifty years were all measures adopted in the pre-digital era for this reason.²⁶³

Beginning in the 1980s, however, the United States took the lead in reshaping the Berne Convention to accommodate computer programs, which many commentators and governments had preferred to view as “electronic information tools”²⁶⁴ subject to more pro-competitive industrial property laws, including patents, unfair competition and hybrid (or *sui generis*) forms of protection.²⁶⁵ By the 1990s, a coalition of content providers concerned about the online copying of movies, music, and software in the new digital environment had persuaded the U.S. government to press for still more far-reaching changes of international copyright and related laws.²⁶⁶ These efforts led to the codification of universal copyright norms in the TRIPS Agreement of 1994²⁶⁷ and to two 1996 World Intellectual Property Organization (“WIPO”) treaties on copyrights and related rights in cyberspace,²⁶⁸ which endowed authors with a bevy of new exclusive rights tailor-made for online transmissions, and which imposed unprecedented obligations on participating governments to prohibit electronic equipment capable of circumventing these rights.²⁶⁹ All of these new norms and obligations, ostensibly adopted to discourage market-destructive copying of literary and artistic works, then became domestic law,²⁷⁰ often with no regard for their impact on science, and sometimes with deliberate disregard of measures adopted to safeguard science and education at the international level.²⁷¹

263. See 17 U.S.C. §§ 102(a), 302, 401(a) (2000) (as amended).

264. See J. H. Reichman, *Electronic Information Tools: The Outer Edge of World Intellectual Property Law*, 24 INT’L REV. INDUS. PROP. & COPYRIGHT L. 446 (1993) [hereinafter Reichman, *Electronic Information Tools*].

265. See, e.g., Pamela Samuelson et al., *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 COLUM. L. REV. 2308 (1994).

266. See generally PETER DRAHOS & JOHN BRAITHWAITE, *INFORMATION FEUDALISM* 107-48 (2002); Gail Evans, *Intellectual Property as a Trade Issue—The Making of the Agreement on Trade-Related Aspects of Intellectual Property Rights*, 18 WORLD COMPETITION L. & ECON. REV. 137 (1994).

267. TRIPS Agreement, *supra* note 63, arts. 9-14.

268. World Intellectual Property Organization (“WIPO”) Copyright Treaty, Dec. 20, 1996, 36 I.L.M. 65 (1996); WIPO Performances and Phonograms Treaty, S. TREATY DOC. NO. 105-17, 36 I.L.M. 76 (1996).

269. See generally Pamela Samuelson, *The U.S. Digital Agenda at WIPO*, 37 VA. J. INT’L L. 369 (1997).

270. See, e.g., The Digital Millennium Copyright Act (“DMCA”), 17 U.S.C. § 1201 (2000); Council Directive 2001/29 of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, 2001 O.J. (L 167) 10 [hereinafter E.C. Directive on Copyright in the Information Society].

271. The DMCA declined to enact concessions on users’ rights that the WIPO Diplomatic Conference and WIPO Copyright Treaty of 1996 had authorized. See also Samuelson, *U.S. Digital Agenda*, *supra* note 269.

At the same time, and as part of the same overall movement, the coalition of content providers that had captured Congress' attention took aim at two closely related areas in which much more than market-destructive copying was actually at stake. The first of these was to validate the uncertain status of standard-form electronic contracts used to regulate online dissemination of works in digital form.²⁷² Because traditional contract and sales laws can be interpreted in ways that limit the kinds of terms that can be imposed through "shrink-wrap" or "click-on" licenses, and the one-sidedness of the resulting "adhesion contracts,"²⁷³ the coalition pushing the high-protectionist digital agenda has sponsored a new uniform law, the Uniform Computer Information Transactions Act ("UCITA")²⁷⁴ to validate such contracts in the form they desire, and it has lobbied state legislatures to adopt it.²⁷⁵

The last major component of the high-protectionists' digital agenda was an attempt by some of the largest database companies to obtain a *sui generis* exclusive property right in non-copyrightable collections of information, even though facts and data had hitherto been off-limits even to international copyright law as reformed under the TRIPS Agreement of 1994.²⁷⁶ These efforts culminated in the European Community's Directive on the Legal Protection of Databases adopted in 1996;²⁷⁷ in a proposed WIPO treaty on the international protection of databases built on the same model, which was barely defeated at the WIPO Diplomatic Conference in December of 1996;²⁷⁸ and in a series of database protection bills that have been introduced in the U.S. Congress that attempt to enact similar measures into U.S. law.²⁷⁹

Most of the developments outlined above resulted from efforts that were not undertaken with science in mind, although publishers who profit from distributing commercialized scientific products promoted some of the changes that appear most threatening for scientific research, especially database protection

272. See Raymond T. Nimmer, *Breaking Barriers: The Relation Between Contract and Intellectual Property*, 13 BERKELEY TECH. L.J. 827, 904-08 (1998).

273. See, e.g., Mark A. Lemley, *Beyond Preemption: the Law and Policy of Intellectual Property Licensing*, 87 CAL. L. REV. 111 (1999) [hereinafter Lemley, *Beyond Preemption*]; Charles R. McManis, *The Privatization or "Shrink-Wrapping" of American Copyright Law*, 87 CAL. L. REV. 173 (1999).

274. UNIF. COMPUTER INFO. TRANSACTIONS ACT ("UCITA") (2001), available at <http://www.ucitaonline.com/> (last visited May 13, 2002) (now adopted in Maryland and Virginia).

275. See generally *Intellectual Property & Contract Law*, *supra* note 16.

276. See TRIPS Agreement, *supra* note 63, at art. 10.2.

277. E.C. Database Directive, *supra* note 17.

278. See Draft Treaty on Intellectual Property Rights in Respect of Databases, WIPO doc. CRNR/DC/6 (1996), available at http://www.wipo.int/eng/diplconf/pdf/6dc_e.pdf; Draft Recommendation, WIPO doc. CRNR/DC/88 (1996), available at <http://www.wipo.int/eng/diplconf/distrib/pdf/88dc.pdf>. The U.S. Patent and Trademark Office and the U.S. Trade Representative initially supported this treaty, together with the Commission of the European Communities. The U.S. position changed following a series of high-level meetings of federal government officials in October and November of 1996, largely in response to a letter from the three presidents of the National Academies to Mickey Kantor, Secretary of Commerce (Oct. 9, 1996) (on file with authors) (expressing serious reservations about the potential effects of such a treaty on scientific research and noting the complete absence of any interagency consultations about this matter or any other public discussion). See generally Reichman & Samuelson, *supra* note 18, at 97-113.

279. See *supra* note 17.

laws. The following sections show that all these measures—whatever their ostensible purpose—have the cumulative effect of shrinking the research commons.

We will first briefly note the impact of selected developments in both federal statutory copyright law and in contract laws at the state level. We then discuss current proposals to confer strong exclusive property rights on non-copy-rightable collections of data, which constitute the clearest and most overt assault on the public domain that has fueled both scientific endeavors and technological innovation in the past.

a. Expanding Copyright Protection of Factual Compilations: The Revolt Against *Feist*

The quest for a new legal regime to protect databases was triggered in part by the U.S. Supreme Court's 1991 decision in *Feist Publications, Inc. v. Rural Telephone Service Co.*,²⁸⁰ which denied copyright protection to the white pages of a telephone directory. As discussed in Part II, that decision reaffirmed the principle that facts and data as such are ineligible for copyright protection as "original and creative works of authorship."²⁸¹ It also limited the scope of copyright protection to any original elements of selection and arrangement that otherwise meet the test of eligibility. Second-comers who developed their own criteria of selection and arrangement could in principle use prior data to make follow-on products without running afoul of the copyright owner's strong exclusive right to prepare derivative works.²⁸² Taken together, these propositions supported the customary and traditional practices of the scientific community and facilitated access to and use of research data.

In recent years, however, judicial concerns about the compilers' inability to appropriate the returns from their investments have induced federal appellate courts to broaden copyright protection of low authorship compilations in ways that significantly deform both the spirit and the letter of *Feist*.²⁸³ At the eligibility stage, so little in the way of original selection and arrangement is now required that the only print media still certain to be excluded from protection

280. 499 U.S. 340, 349-51, 359-60 (1991).

281. 17 U.S.C. §§ 102(a), 102(b), 103 (2000). See discussion of *Feist supra* at text accompanying notes 88-89. See also *Key Publ'ns, Inc. v. Chinatown, Today Publ'g Enter., Inc.*, 945 F.2d 509, 514 (2d Cir. 1991) ("thin" copyright doctrine).

282. 17 U.S.C. §§ 101, 102, 103, 106(2) (2000); *Feist*, 499 U.S. at 354 (stressing adverse effects on free flow of information by creating "monopol[ies] in public domain materials"); see also Jane C. Ginsburg, *No "Sweat?" Copyright and Other Protection of Works of Information After Feist v. Rural Telephone*, 92 COLUM. L. REV. 338, 339 (1992) [hereinafter Ginsburg, "No Sweat?"]; Jessica Litman, *After Feist*, 17 U. DAYTON L. REV. 607, 609 (1992).

283. See, e.g., *CDN, Inc., v. Kapes*, 197 F.3d 1256, 1259-60 (9th Cir. 1999); *CCC Info. Servs., Inc. v. Maclean Hunter Mkt. Reports, Inc.*, 44 F.3d 61, 65 (2d Cir. 1994) (stressing low threshold of eligibility); cf. *Am. Dental Ass'n v. Delta Dental Plans Ass'n*, 126 F.3d 977 (7th Cir. 1997) (taxonomies of dental procedures not excluded subject matter); Justin Hughes, *Created Facts – or the Occasional Protection of Ideas, Names and Facts in Copyright Law* (forthcoming 2003, on file with authors).

are the white pages of telephone directories.²⁸⁴

More tellingly, the courts have increasingly perceived the eligibility criteria of selection and arrangement as pervading the data themselves, in order to restrain second-comers from using pre-existing data sets to perform operations that are functionally equivalent to those of an initial compiler.²⁸⁵ In the Second Circuit, for example, a competitor could not assess used car values by the same technical means employed in a first-comer's copyrightable compilation, even if those means turned out to be particularly efficient, and even if the second-comer combined the protected valuations with those of another rating system in an averaged set of values.²⁸⁶ Similarly, the Ninth Circuit prevented even the use of a small amount of data from a copyrighted compilation that was essential to achieving a functional result.²⁸⁷

Copyright law provides a very long term of protection, and it generally endows authors with strong rights to control follow-on applications of the protectible contents of their works.²⁸⁸ Stretching copyright law to cover algorithms and aggregates of facts (and even so-called "soft" or subjective ideas), as these recent decisions have done, conflates the idea-expression dichotomy and indirectly extends protection to facts as such.²⁸⁹

Opponents of *sui generis* database protection in the United States cite these and other cases as evidence that no *sui generis* database protection law is needed.²⁹⁰ In reality, these cases suggest that, in the absence of a suitable minimalist regime of database protection to alleviate the risk of market failure without impoverishing the public domain,²⁹¹ courts tend to convert copyright law into a roving unfair competition law that can protect both factual and functional matter, including algorithms, for very long periods of time and that could create formidable barriers to entry. This tendency, however, ignores the historical

284. See, e.g., *Warren Publ'g Inc. v. Microdos Data Corp.*, 115 F.3d 1509, 1518-19 (11th Cir. 1997) (en banc); *Bellsouth Adver. & Publ'g Corp. v. Donnelley Info. Publ'g, Inc.*, 999 F.2d 1436, 1446 (11th Cir. 1993) (en banc).

285. See, e.g., Ginsburg, *U.S. Initiatives*, *supra* note 18, at 57-61 (explaining protection of subjective criteria of selection and arrangement as distinct from objective criteria, as set out in *CCC Info. Servs.*, 44 F.3d at 71); see also Dennis S. Karjala, *Copyright in Electronic Maps*, 35 JURIMETRICS J. 395, 408-11 (1995).

286. *CCC Info. Servs.*, 44 F.3d at 71-74. But see *Baker v. Selden*, 101 U.S. 99 (1879) (denying copyright protection to functional aspects of literary works).

287. *CDN, Inc.*, 197 F.3d at 1256. See generally Hughes, *supra* note 283.

288. 17 U.S.C. §§ 102(a), 103, 106(2), 302-304 (2000); *Eldred v. Reno*, 239 F.3d 372 (D.C. Cir. 2001), *cert. granted sub nom*, *Eldred v. Ashcroft*, 534 U.S. 1062, and *cert. amended*, 534 U.S. 1160 (2002); see also Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989 (1997) [hereinafter Lemley, *Improvement*].

289. See, e.g., Alan L. Durham, Note, *Speaking of the World: Fact, Opinion and the Originality Standard of Copyright*, 33 ARIZ. ST. L.J. 791, 838-42 (2001).

290. See *supra* note 20; see also Johathan Band & Makoto Kono, *The Database Protection Debate in the 106th Congress*, 62 OHIO ST. L.J. 869 (2001).

291. See Reichman & Samuelson, *supra* note 18, at 137-150 (proposing minimalist regimes of database protection).

limits of copyright protection in defiance of well-established Supreme Court precedent,²⁹² and ultimately jeopardizes access to the research commons.

b. The DMCA: An Exclusive Right to Access Minimally Copyrightable Compilations of Data?

With regard to copyrightable compilations of data distributed online, amendments to the Copyright Act of 1976, known as the Digital Millennium Copyright Act of 1998 (“DMCA”),²⁹³ may have greatly reduced the traditional safeguards surrounding research uses of factual works. Technically, section 1201(a) establishes a right to prevent the direct circumvention of any electronic fencing devices that a content provider may have employed to control access to a copyrighted work.²⁹⁴ Section 1201(b) then perfects the scheme by exposing manufacturers and suppliers of equipment capable of circumventing electronic fencing devices to liability for copyright infringement when such equipment can be used to violate the exclusive rights traditionally held by copyright owners.²⁹⁵

In enacting these provisions, Congress seems to have detached the prohibition against gaining unauthorized direct access to electronically fenced works under section 1201(a) from the balance of public and private interests otherwise established in the Copyright Act of 1976.²⁹⁶ As Professor Jane Ginsburg interprets this provision, a violation of section 1201(a) is not an “infringement of copyright” because it attracts a separate set of distinct remedies set out in section 1203,²⁹⁷ and because it constitutes “a new violation” for which those remedies are provided.²⁹⁸ On this reading, unlawful access is not subject to the traditional defenses and immunities of the copyright law, and one is “not . . . permitted to circumvent the access controls, even to perform acts that are lawful under the Copyright Act,”²⁹⁹ including presumably the user’s right to extract unprotectible facts and ideas or to invoke the fair use defense.³⁰⁰ On the

292. *Baker*, 101 U.S. at 99; J. H. Reichman, *Computer Programs as Applied Scientific Know-How: Implications of Copyright Protection for Commercialized University Research*, 42 VAND. L. REV. 639, 693-94, n.288 (1989) [hereinafter Reichman, *Computer Programs*] (on the deeper meaning of *Baker*).

293. Pub. L. No. 105-304, 112 Stat. 2860 (1998) (codified at 17 U.S.C. § 1201 (2000)).

294. “No person shall circumvent a technological measure that effectively controls access to a work protected under this title.” 17 U.S.C. § 1201(a) (2000).

295. 17 U.S.C. § 1201(b).

Manufacture and distribution of post-access circumvention devices and services are prohibited only if they are “primarily designed” or have “only limited commercially significant purpose or use other than” to circumvent “protection afforded by a technological measure that effectively protects a right of a copyright owner under this title in a work or a portion thereof” or if they are marked “as circumvention devices.”

Ginsburg, *U.S. Initiatives*, *supra* note 18, at 65 (interpreting 17 U.S.C. § 1201(b)(1)(A)-(C)).

296. See Ginsburg, *U.S. Initiatives*, *supra* note 18, at 62-65

297. 17 U.S.C. § 1203 (2000).

298. Ginsburg, *U.S. Initiatives*, *supra* note 18, at 63-64.

299. *Id.* at 62.

300. See *id.* at 62-64. It remains possible that a court could interpret around these provisions to reach a fair use defense in such a case, even though section 1201(c) seems to restrict that defense to post-access uses. See 17 U.S.C. § 1201(c); Ginsburg, *U.S. Initiatives*, *supra* note 18, at 63-64. Obstructing access to non-copyrightable components might also attract the misuse defense in appropriate circumstances. Cf. Frischmann & Moylan, *supra* note 106.

contrary, “Congress may in effect have extended copyright to cover ‘use’ of works of authorship, including minimally original databases . . . because ‘access’ is a prerequisite to ‘use,’ [and] by controlling the former, the copyright owner may well end up preventing or conditioning the latter.”³⁰¹

While the precise contours of these provisions remain to be worked out in future judicial decisions,³⁰² they could potentiate the ability of both publishers and scientists to protect online collections of data that were heretofore unprotectible in print media. If, for example, a database provider combined the non-copyrightable collection of data with a nominally copyrightable component, such as an analytical explanation of how the data were compiled, the “fig leaf” copyrightable component might suffice to trigger the “no direct access” provisions of section 1201(a).³⁰³ In that event, later scientific researchers could not circumvent the electronic fence in order to extract or use the non-copyrightable data, even for nonprofit scientific research, because section 1201(a) does not recognize the normal exceptions to copyright protection that would allow such use and scientific research is not one of the few very limited exceptions that were codified in section 1201(d)-(j).³⁰⁴

Later researchers would thus have to acquire lawful access to the electronically fenced database under section 1201(a) and then attempt to extract the non-copyrightable data for nonprofit research purposes under section 1201(b) which does, in principle, recognize the traditional users’ defenses as well as the privileges and immunities codified in sections 107-122 of the Copyright Act of 1976.³⁰⁵ Even here, however, later scientists could discover that the technical devices they had used to extract non-protectible data from minimally copyrightable databases independently violated section 1201(b) of the DMCA

301. Ginsburg, *U.S. Initiatives*, *supra* note 18, at 63.

302. For the moment, they have withstood attack on constitutional grounds. *See, e.g.,* A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004, 1014-18 (9th Cir. 2001). *See generally* DIGITAL DILEMMA, *supra* note 14; JESSICA LITMAN, DIGITAL COPYRIGHT: PROTECTING INTELLECTUAL PROPERTY ON THE INTERNET (2001); Pamela Samuelson & Randall Davis, *The Digital Dilemma: A Perspective on Intellectual Property in the Information Age*, available at <http://www.sims.berkeley.edu/~pam/papers/digdilsyn.pdf> (last visited Nov. 15, 2002).

303. *See* Ginsburg, *U.S. Initiatives*, *supra* note 18, at 63 (finding that “the copyrightable fig leaf a database producer affixes to an otherwise unprotectible work could, as a practical matter, obscure the public domain nakedness of the compiled information”).

304. *See* 17 U.S.C. § 1201(d)-(j) (2000). *But see* discussion *supra* note 300. In principle, scientific bodies could petition for an exemption from § 1201(a) as adversely affected users, or they could seek to argue that non-infringing uses of copyrightable databases should fall within a class of adversely affected uses, but the Copyright Office has so far resisted this approach. *See* 17 U.S.C. § 1201(a)(1)(C), (D); Ginsburg, *U.S. Initiatives*, *supra* note 18, at 64 nn.40, 41.

305. Ginsburg, *U.S. Initiatives*, *supra* note 18, at 65-67. Assuming that the end-user cannot circumvent the electronic fence without recourse to some technical device, much would depend on the extent to which any such device could readily be used for infringing as well as non-infringing uses.

Section 1201(b) thus seems to lead to an impasse: it is permissible to circumvent anticopying controls in order to make noninfringing use, but the software or device needed to engage in the circumvention cannot be disseminated because it can all too easily be put to infringing use. *Id.* at 66-67 (exploring the possibility of requiring content providers to identify nonprotectible components).

because those devices were otherwise capable of substantial infringing uses.³⁰⁶ In practice, moreover, the posterior scientists' theoretical opportunity to extract non-copyrightable data by technical devices that did not violate section 1201(b) could already have been compromised by the electronic contracts these scientists will have accepted in order to gain lawful access to the online database in the first place to avoid the crushing power of section 1201(a). In that event, the scientists would almost certainly have waived any user rights they had retained under section 1201(b), unless the electronic contracts themselves became unenforceable on one ground or another, as discussed below.³⁰⁷

In effect, the DMCA allows copyright owners to surround their collections of data with technological fences and electronic identity marks buttressed by encryption and other digital controls that force would-be users to enter the system through an electronic gateway.³⁰⁸ To pass through the gateway, users must accede to non-negotiable electronic contracts, which impose the copyright owner's terms and conditions without regard to the traditional defenses and statutory immunities of copyright law.³⁰⁹

The DMCA indirectly recognized the potential conflict between proprietors and users of ineligible material, such as facts and data, that section 1201(a) of the statute could thus trigger, and it empowered the Copyright Office, which reports to the Librarian of Congress, to exempt categories of users whose activities might be adversely affected.³¹⁰ While representatives of the educational and library communities petitioned for relief on various grounds, including the need of researchers to access and use non-copyrightable facts and ideas transmitted online, the authorities have so far declined to act.³¹¹

It is too soon to know how far owners of copyrightable compilations can push this so-called "right of access"³¹² at the expense of research, competition, and free speech without incurring resistance based on the misuse doctrine of copyright law, the public policy and unconscionability doctrines of state contract laws, and First Amendment concerns that have in the past limited copyright protection of factual works.³¹³ For the foreseeable future, nonetheless, the

306. See 17 U.S.C. § 1201(b)(c).

307. See Reichman & Franklin, *supra* note 16.

308. Cohen, *Lochner in Cyberspace*, *supra* note 97; see also 17 U.S.C. § 1202 (prohibiting removal of copyright management information).

309. See, e.g., McManis, *supra* note 273. See generally GUIBAULT, *supra* note 95, at 291-304. Attempts to bypass these electronic barriers in the name of pre-existing legal defenses then constitute either a violation of the access right under section 1201(a), which could impede third parties from raising even the well-established traditional defenses to an action for infringement, or an independent basis for infringement under § 1201(b). See *supra* notes 296-301, 306 and accompanying text.

310. 17 U.S.C. § 1201(a)(1)(C)-(D).

311. See *supra* note 304; *Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies*, 37 C.F.R. § 201 (2000);

312. See Robert A. Kreiss, *Accessibility and Commercialization in Copyright Theory*, 43 U.C.L.A. L. REV. 1, 32-34 (1995) (discussing thesis that copyrighted works should be accessible).

313. Lemley, *Beyond Preemption*, *supra* note 273; Reichman & Franklin, *supra* note 16, at 929-53 (proposing and applying doctrine of "public interest unconscionability" as functional equivalent of misuse doctrine).

DMCA empowers owners of copyrightable collections of facts to contractually limit online access to the pre-existing public domain in ways that contrast drastically with the traditional availability of factual contents in printed works.

c. One-Sided Electronic Licensing Contracts

Data published in print media traditionally entered the public domain under the classical intellectual property regime described above. Further ensuring that result is an ancillary copyright doctrine, known as “exhaustion” or “first sale doctrine,” which limits the authors’ powers to control the uses that third parties can make of copyrighted literary works distributed to the public in hard copies.³¹⁴

Under this doctrine, the copyright owner may extract a profit from the first sale of the copy embodying an original and protectible compilation of data, but cannot prevent a purchaser from reselling that physical copy or from using it in any way the latter deems fit, say, for research purposes, unless such uses amount to infringing reproductions, adaptations, or performances of the expressive components of the copyrighted compilation.³¹⁵ In effect, copyright law not only made it difficult to protect compilations of data as such, but it denied authors any exclusive right to control the use of a protected work once it had been distributed to the public in hard copies.³¹⁶

The first sale doctrine thus complements and perfects the other science-friendly provisions described above, unless individual scientists, libraries, or scientific entities were to contractually waive their rights to use copies of purchased works in the manner described. Such contractual waivers always remain theoretically possible, and publishers have increasingly pressed them upon the scientific and educational communities in the online environment for reasons discussed below.

Nevertheless, it was not generally feasible to impose such waivers against scientists who bought scientific works distributed to the public in hard copies, and even when attempts to do so were made, such contracts could not bind subsequent purchasers of the copies in question. The upshot was that, precisely because authors and publishers could not rely on contractual agreements, they depended on the default rules of copyright law, which are binding against the world. These default rules, in turn, impose legislatively enacted “contracts,” which balance public and private interests by, for example, defining the uses that libraries can make of their copies,³¹⁷ immunizing certain protected uses for

314. 17 U.S.C. § 109(a) (2000); *see also* § 117 (allowing owner of copyrighted computer program to use it on any computer).

315. *See* 17 U.S.C. § 106 (delineating exclusive rights to reproduce, adapt, publicly perform, distribute, and display copyrighted works, but conferring no exclusive right to use); § 109(a) (defining right of owner to dispose of physical copy of protected work); Ralph S. Brown, *Eligibility for Copyright Protection: A Search for Principled Standards*, 70 MINN. L. REV. 579, 588-89 (1985) (stressing importance of denial of exclusive use).

316. *Accord* Ginsburg, *U.S. Initiatives*, *supra* note 18, at 63.

317. 17 U.S.C. § 108.

educational purposes,³¹⁸ and further allowing a set of fair uses that scientists and other researchers can invoke.³¹⁹

i. Restoring the Power of the “Two-Party” Deal

Against this background, online delivery of both copyrightable and non-copyrightable productions possesses the inherent capabilities of changing the pre-existing relationship between authors and readers or between content providers and users. As previously discussed, by placing a collection of minimally copyrightable data online and surrounding it with technological fencing devices, publishers can condition access to the database on the would-be user’s acquiescing to the terms and conditions of the former’s “click-on,” standard-form, non-negotiable contract (known as a contract of adhesion).³²⁰ To this end, highly restrictive digital rights management technologies (“DRM”) are being developed that include hardware and software based “trusted systems,” online database access controls, and increasingly effective forms of encryption.³²¹

The power to control online access that digital rights management technologies confers is, moreover, conceptually and empirically independent of statutory intellectual property rights, which makes it of capital importance for the theses discussed in this article. It means that even if a given compilation of data lacked any copyrightable “fig leaf” whatsoever, so that it could not trigger the so-called “access right” that section 1201(a) of the DMCA otherwise provides,³²² the electronic contract accepted at the gateway to the provider’s electronic fence may itself enable him to control all the uses of the non-copyrightable data, which would technically enter the public domain. So long as third parties cannot feasibly acquire the data in question except by individually accepting the online provider’s “click-on” licensing restrictions, online delivery can solve most of the problems that the printing press created for authors by enabling them contractually to restrict the use of productions made available to the public, whether copyrightable or not, and in this sense it restores the “power of the two-party

318. See 17 U.S.C. § 110(1), (2).

319. 17 U.S.C. § 107; *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 573-96 (1994); *Harper & Row Publ’s, Inc. v. Nation Enters.*, 471 U.S. 539, 569 (1985).

320. See, e.g., 17 U.S.C. §§ 1201(a), 1202; *McManis*, *supra* note 273; see also *GUIBAULT*, *supra* note 95, at 3.

321. See, e.g., Julie E. Cohen, *A Right to Read Anonymously: A Closer Look at “Copyright Management” in Cyberspace*, 28 *CONN. L. REV.* 981, 983-89 (1996) (discussing technologies that copyright owners may employ to monitor and control access to information); Jessica Litman, *The Exclusive Right to Read*, 13 *CARDOZO ARTS & ENT. L.J.* 29 (1994); Mark Stefik, *Shifting the Possible: How Trusted Systems and Digital Property Rights Challenge Us to Rethink Digital Publishing*, 12 *BERKELEY TECH. L.J.* 137 (1997); see also PETER ECKERSLEY, *VIRTUAL MARKETS FOR VIRTUAL GOODS: AN ALTERNATIVE CONCEPTION OF DIGITAL COPYRIGHT*, *Intell. Prop. Res. Inst. of Australia Working Paper* (2002), available at <http://www.cs.mu.oz.au/~pde/writing/virtualmarkets.pdf>; Dan L. Burk & Julie Cohen, *Fair Use Infrastructure for Copyright Management Systems*, 15 *HARV. J.L. & TECH.* 41-83 (2001); Pamela Samuelson, *Intellectual Property and the Digital Economy: Why the Anti-Circumvention Regulations Need to be Revised*, 14 *BERKELEY TECH. L.J.* 519 (1999).

322. The presence of some original copyrightable expression remains a necessary prerequisite to triggering the technical protection measures of the DMCA, which to that extent remain copyright dependent. See 17 U.S.C. §§ 102(a)(b), 103 (2000); Ginsburg, *U.S. Initiatives*, *supra* note 18, at 63.

deal” that publishers lost in the sixteenth century.³²³

Because electronic contracts are enforceable in state courts, they provide private rights of action that tend to either substitute for or override statutory intellectual property rights. Electronic contracts become substitutes for intellectual property rights to the extent that they make it infeasible for third parties to obtain publicly disclosed but electronically fenced data without incurring contractual liability for damages. They may override statutory intellectual property rights by, for example, forbidding the uses that libraries could otherwise make of a scientific work under federal copyright law, or by prohibiting follow-on applications or the reverse engineering of a computer program that both federal copyright law and state trade secret law would otherwise permit.³²⁴

To the extent that those who draft electronic contracts are allowed to impose terms and conditions that ignore the goals and policies of the federal intellectual property system, they could establish “privately legislated intellectual property rights” unencumbered by concessions to the public interest.³²⁵ By the same token, a privately generated database protected by technical devices and electronic adhesion contracts is subject to no federally imposed duration clause and accordingly will never lapse into the public domain.

ii. The Proposed Uniform Computerized Information Transactions Act (“UCITA”)

Whether state courts should enforce electronic contracts—especially the non-negotiable, standard-form “click-on” and “shrink-wrap” contracts—remains an open and controversial question.³²⁶ Besides technical obstacles to formation sounding in general contracts law, commentators argue that courts may deem such contracts unenforceable under the public policy defense of state contracts law, under the preemption doctrine that supports the integrity of the federal intellectual property system, or under some combination of the two.³²⁷ In this regard, the doctrine of unconscionability, spawned by the Uniform Commercial Code,³²⁸ could be expanded to encompass a concept of “public interest unconscionability,” which in effect would endow state courts with a

323. See Reichman & Franklin, *supra* note 16, at 897-99.

324. See, e.g., Rochelle Cooper Dreyfuss, *Do You Want to Know a Trade Secret? How Article 2B Will Make Licensing Trade Secrets Easier (But Innovation More Difficult)*, 87 CAL. L. REV. 191 (1999). But see, e.g., 17 U.S.C. § 301 (2000); Dennis S. Karjala, *Federal Preemption of Shrinkwrap and On-Line Licenses*, 22 U. DAYTON L. REV. 511 (1997); David A. Rice, *Public Goods, Private Contract and Public Policy: Federal Preemption of Software License Prohibitions Against Reverse Engineering*, 53 U. PITT. L. REV. 543 (1992).

325. See generally Reichman & Franklin, *supra* note 16, at 929-53 (proposing misuse of contracts doctrine to regulate privately legislated intellectual property rights that disrupt the public-private balance of the federal system without reasonable economic justification).

326. See, e.g., *Hill v. Gateway 2000, Inc.*, 105 F.3d 1147, 1148 (7th Cir. 1997), *cert. denied*, 522 U.S. 808 (1997); *ProCD v. Zeidenberg*, 86 F.3d 1447 (7th Cir. 1996). But see Alexander M. Meiklejohn, *Castles in the Air: Blanket Assent and the Revision of Article 2*, 51 WASH. & LEE L. REV. 599, 603 (1994).

327. See generally *Intellectual Property and Contract Law*, *supra* note 16

328. U.C.C. § 2-302 (2002).

“misuse of contracts” concept to parallel and dovetail with the doctrines of “misuse of intellectual property rights.”³²⁹

In practice, however, courts appear reluctant to exercise such powers even when their right to do so is clear. The most recent line of cases, led by the Seventh Circuit’s opinion in *Pro-CD v. Zeidenberg*,³³⁰ has tended to validate electronic contracts of adhesion in the name of “freedom of contract.” In this same vein, the National Council of Commissioners for Uniform State Law (“NCCUSL”) has proposed a Uniform Computer Information Transactions Act (“UCITA”), which, if state legislatures enacted it, would broadly validate such contracts and largely immunize them from legal challenge.³³¹

For example, UCITA permits vendors of information products to define virtually every transaction as a “license” rather than a “sale,” and it tolerates perpetual licenses.³³² It could thus override the first-sale doctrine of copyright law and any analogous doctrine that might be embodied in the proposed database protection laws discussed below.

The proposed uniform law would then proceed to broadly validate mass market “click-on” and “shrink-wrap” licenses that impose all the provisions vendors could hope for, with little regard for the interests of scientific and educational users, or the public in general.³³³ It would permit vendors to add further, non-negotiated conditions to the perpetual licensing agreement even after the product had been paid for, and in case of dispute, it would permit vendors to block recalcitrant “licensees” who too vigorously complained about either the product or the terms and conditions that accompany it from further accessing or using the information it made available.³³⁴

A detailed analysis of UCITA’s provisions is beyond the scope of this study. Suffice it to say, however, that its less-than-transparent drafting process so favored the interests of sellers of software and other information products at the expense of consumers and users generally that a coalition of sixteen state attorneys general vigorously opposed its adoption, and the American Law Institute withdrew its co-sponsorship of the original project. Nonetheless, two states—Maryland and Virginia—have adopted non-uniform versions of

329. See Reichman & Franklin, *supra* note 16, at 927-37 (“Validating Non-negotiable Terms that Respect the Balance of Public and Private Interests”); see also Niva Elkin-Koren, *A Public-Regarding Approach to Contracting Over Copyright*, in EXPANDING THE BOUNDARIES OF IP, *supra* note 13, at 191-222.

330. 86 F.3d at 1447; *Intellectual Property and Contract Law*, *supra* note 16.

331. UNIF. COMPUTER INFO. TRANSACTIONS ACT (UCITA) (2001), available at <http://www.ucitaonline.com/> (last visited Feb. 14, 2003). At the time of this writing, the NCCUSL had just completed some revisions to UCITA that reportedly address some of the most excessive features of the proposed act that heavily favor the rights of vendors over licensees. Nonetheless, the American Bar Association has refused to endorse it.

332. UCITA § 308.

333. UCITA § 209. See generally Reichman & Franklin, *supra* note 16, at 938-65.

334. See, e.g., Julie E. Cohen, *Copyright and the Jurisprudence of Self-Help*, 13 BERKELEY TECH. L. J. 1089 (1998).

UCITA,³³⁵ and major software and information industry firms continue to lobby assiduously for its enactment by other state legislatures.

If present trends continue unabated, privately generated information products delivered online—including databases and computer software—may be kept under a kind of perpetual, mass market trade secret protection, subject to no reverse engineering efforts or public-interest uses that are not expressly sanctioned by licensing agreements. Contractual rights of this kind, backed by a one-sided regulatory framework, such as UCITA, could conceivably produce an even higher level of protection than that available from some future federal database right subject to statutory public-interest exceptions. The most powerful proprietary cocktail of all, however, would probably emerge from a combination of a strong federal database right with UCITA-backed contracts of adhesion.

d. New Exclusive Property Rights in Non-Copyrightable Collections of Data

The challenge of protecting commercially valuable collections of information that fail to meet the technical eligibility requirements of copyright law poses a hard problem that has existed in one form or another for two centuries,³³⁶ and at least three different approaches have emerged over time.³³⁷ One solution would allow a domestic copyright law to accommodate low authorship literary productions, with some adjustments to the bundle of rights at the margins.³³⁸ A second approach, adopted in the Nordic countries, would enact a short term *sui generis* regime, built on a distinctly copyright-like model that would protect catalogues, directories, and tables of data against wholesale duplication, without conferring on proprietors any exclusive adaptation right like that afforded to authors of true literary and artistic works.³³⁹ A third approach, experimented with at different times and to varying degrees in different countries, including the United States, would protect compilers of information against wholesale duplication of their products under different theories rooted in the misappropriation branch of unfair competition law.³⁴⁰

335. Robert A. Hillman & Jeffrey J. Rachlinski, *Standard-Form Contracting in the Electronic Age*, 77 N.Y.U. L. REV. 429, 491 n.314 (2002).

336. See, e.g., Ginsburg, *U.S. Initiatives*, *supra* note 18, at 55.

337. This section is based on J. H. Reichman, *Database Protection in a Global Economy*, *supra* note 18, at 455, 459-84.

338. See, e.g., *Desktop Mktg Sys. Pty. Ltd. v. Telstra Corp. Ltd.* (2002) 192 A.L.R. 433 (Austl.) (holding that white pages of telephone directory qualify for copyright protection under “sweat of the brow” theory). See generally Baron, *supra* note 78; Denicola, *supra* note 78; Ginsburg, “No Sweat?,” *supra* note 282; Ginsburg, *Commercial Value*, *supra* note 78.

339. See, e.g., Gunnar W. G. Karnell, *The Nordic Catalogue Rule*, in PROTECTING WORKS OF FACT 67 (Egbert J. Dommering & P. Bernt Hugenholtz eds., 1991).

340. *Int’l News Serv., Inc. v. Assoc. Press*, 248 U.S. 215 (1918); *Nat’l Basketball Ass’n v. Motorola Inc.*, 105 F.3d 841 (2d Cir. 1997); Jane C. Ginsburg, *Copyright, Common Law and Sui Generis Protection of Databases in the United States and Abroad*, 66 U. CIN. L. REV. 151 (1997); see also Gordon, *supra* note 152; Jason R. Boyarski, *The Heist of Feist: Protection for Collections of Information and the Possible Federalization of “Hot News,”* 21 CARDOZO L. REV. 871 (1999); Brian F. Fitzgerald, *Protecting*

What changed in the 1990s was the convergence of digital and telecommunications networks, which potentiated the role of electronic databases in the information economy generally, and which made scientific databases in particular into agents of technological innovation whose economic potential may eventually outstrip that accruing from the patent system.³⁴¹ Notwithstanding the robust appearance of the present day database industry under free market conditions,³⁴² analysts asked whether inadequate investment in complex digital databases would not inevitably hinder that industry's long-term growth prospects if free-riding second-comers could rapidly appropriate the contents of successful new products without contributing to their costs of development and maintenance over time. In other words, if copyright, contract law, DRM technologies, residual unfair competition laws, and various protective business practices inadequately filled a gap in the law, then regulatory action to enhance investment might be justified.³⁴³ This utilitarian rationale, however, raised new and still largely unaddressed questions about the unintended social costs likely to ensue if intellectual property rights were injudiciously bestowed upon the raw materials of the information economy in general and on the building blocks of scientific research in particular.³⁴⁴

Any serious effort to find an appropriate *sui generis* solution to the question of database protection should have engendered an investigation of the comparative economic advantages and disadvantages of regimes based on exclusive property rights as distinct from regimes based on unfair competition laws and other forms of liability rules.³⁴⁵ This investigation should also have taken account of larger questions about the varying impacts of different legal regimes on freedom of speech and the conditions of democratic discourse, which, in the United States at least, are of primary constitutional importance.³⁴⁶ Instead, the Commission of the European Community cut the inquiry short by adopting the

Informational Products (Including Databases) Through Unjust Enrichment Law: An Australian Perspective, 1998 E.I.P.R. 244 (1998).

341. See, e.g., Reichman & Uhler, *Database Protection*, *supra* note 4; cf. Reichman, *Electronic Information Tools*, *supra* note 264. The emergence of digitally networked environments has generated a host of new value-added services and products, and appreciably increased the importance of this segment of the database market. David Fewer, *Constitutionalizing Copyright: Freedom of Expression and the Limits of Copyright in Canada*, 55 U. TORONTO FAC. L. REV. 175 (1997); see also Hunsucker, *supra* note 239.

342. See Fewer, *supra* note 341 (case of Canada); Maurer, *Across Two Worlds*, *supra* note 34. See generally A QUESTION OF BALANCE, *supra* note 1.

343. See, e.g., Hunsucker, *supra* note 239; *Information Antipiracy Hearings*, *supra* note 239.

344. See BITS OF POWER, *supra* note 1; Reichman & Samuelson, *supra* note 18, at 113-37.

345. See Reichman, *Of Green Tulips*, *supra* note 164; *supra* note 151 (defining liability rules).

346. For example, the constitutional foundations of United States copyright law have always rested on a clear and sharp distinction between facts and ideas that were freely available to all and the author's expression of facts and ideas, which could not be copied. 17 U.S.C. § 102(b) (2000). Allowing exclusive property rights to cover aggregates of data and information that had been previously unprotected must sooner or later pose fundamental constitutional questions for countries that take freedom of speech seriously, questions that a creative use of liability rules might altogether avoid. See generally Benkler, *Constitutional Bounds*, *supra* note 91; Hamilton, *supra* note 91; Paul J. Heald, *The Extraction/Duplication Dichotomy: Constitutional Line-Drawing in the Database Debate*, 62 OHIO ST. L.J. 933 (2001).

Directive on the Legal Protection of Databases in 1996.³⁴⁷ This Directive requires all E.U. member countries (and affiliated states) to pass laws that confer a hybrid exclusive property right on publishers who make substantial investments in non-copyrightable compilations of facts and information.³⁴⁸

i. The E.C. Database Directive in Brief

The hybrid exclusive right that the European Commission ultimately crafted in its Directive on the Legal Protection of Databases does not resemble any pre-existing intellectual property regime. It protects any collection of data, information, or other materials that is arranged in a systematic or methodological way, provided that it is individually accessible by electronic or other means.³⁴⁹ To become eligible for protection, the database producer must demonstrate a “substantial investment,” as measured in either qualitative or quantitative terms,³⁵⁰ which leaves the courts to develop this criterion with little guidance from the legislative history.³⁵¹ The drafters explicitly recognized that the qualifying investment may consist of no more than simply verifying or maintaining the database.³⁵²

In return for this investment, the compiler obtains exclusive rights to extract or reutilize all or a substantial part of the contents of the protected database.³⁵³ The exclusive extraction right pertains to any transfer in any form of all or a substantial part of the contents of a protected database;³⁵⁴ the exclusive reutilization right, by contrast, covers only the making available to the public of all or a substantial part of the same database, typically by incorporation of those data into another database.³⁵⁵ In every case, the first-comer obtains an exclusive right to control uses of collected data as such, as well as a powerful adaptation (or derivative work) right along the lines that copyright law bestows on “original works of authorship,”³⁵⁶ even though such a right is alien to the protection of investment under existing unfair competition laws.³⁵⁷ In a recent interpretation of this provision, a United Kingdom court vigorously enforced this right to control follow-on applications of an original database against a value-adding sec-

347. E.C. Database Directive, *supra* note 17.

348. *Id.* at arts. 1(2), 7(1).

349. *Id.* at art. 1(2).

350. *Id.* at art. 7(1).

351. The Recitals, and some implementing laws, state that the “substantial investment” may entail financial, material, or human resources. *See, e.g., id.*, Recital 40; *see also* Ginsburg, *U.S. Initiatives*, *supra* note 18, at 69 n.55.

352. E.C. Database Directive, *supra* note 17, at art. 7(1).

353. *Id.* at arts. 7(1)-(2).

354. *Id.* at art. 7(2)(a). For example, any transfer of all or a substantial part of a paper or disc will trigger this clause. “Indeed, extraction also occurs if the user simply downloads a ‘substantial part’ . . . to ‘RAM’ to view on a screen since ‘extraction’ . . . shall mean the permanent *or temporary* transfer.” Ginsburg, *U.S. Initiatives*, *supra* note 18, at 69.

355. E.C. Database Directive, *supra* note 17, at art. 7(2)(b).

356. *See, e.g.*, 17 U.S.C. §§ 101 (“derivative works”), 103, 106(2).

357. *See* Reichman & Samuelson, *supra* note 18, at 137-51.

ond-comer.³⁵⁸ It took this position even though the proprietor was the sole source of the data in question and there was no feasible way to generate them by independent means.³⁵⁹

The Directive contains no provision expressly regulating the collections of information that member governments themselves produce. This lacuna leaves European governments that generate data free to exercise either copyrights³⁶⁰ or *sui generis* rights in their own productions in keeping with their respective domestic policies. This result contrasts sharply with the situation in the United States, where the government cannot claim intellectual property rights in the data it generates and must normally make such data available to the public for no more than a cost-of-delivery fee.³⁶¹

The Directive provides no mandatory public-interest exceptions comparable to those recognized under domestic and international copyright laws. An optional, but ambiguous, exception concerning “illustrations for teaching or scientific research” applies to extractions but not reutilization.³⁶² This provision would prevent a nonprofit scientist from incorporating an extract taken from a protected database into a new and different compilation.³⁶³

The Directive’s *sui generis* regime exempts from liability anyone who extracts or reuses an insubstantial part of a protected database, and this exception may not be overridden by contract.³⁶⁴ However, such a user bears the risk of accurately drawing the line between a substantial and an insubstantial part, and any repeated or systematic uses of even an insubstantial part will forfeit this exemption.³⁶⁵ Judicial interpretation has so far taken a restrictive view of this exemption, and one cannot effectively make unauthorized extractions or uses of an insubstantial part of any protected database without serious risk of triggering an action for infringement.³⁶⁶

Qualifying databases are nominally protected for a fifteen-year period.³⁶⁷ In reality, each new substantial investment in a protected database, such as the

358. British Horseracing Bd. Ltd. v. William Hill Org. Ltd., 2001 E.W.C.A. Civ. 1268 (Eng. C.A.).

359. *Id.*

360. See E.C. Database Directive, *supra* note 17, at ch. I, arts. 1-6; Berne Convention, *supra* note 109, at art. 2(4) (leaving this to discretion of governments).

361. 17 U.S.C. § 105 (2000).

362. E.C. Database Directive, *supra* note 17, at arts. 9, 9(b). This exception requires attribution and must not exceed an amount “justified by the noncommercial purpose to be achieved”); see also *id.* at art. 9(a) (broader exception for private use extractions from hard copies).

363. See, e.g., Reichman & Uhler, *Database Protection*, *supra* note 4, at 803; accord Ginsburg, *U.S. Initiatives*, *supra* note 18, at 69-70. This provision may be open to a more flexible interpretation, and some member countries, notably the Nordic countries, have implemented a broader version. Other countries, notably France, Italy, and Greece, have simply ignored this exception altogether, which defeats the Commission’s supposed concerns to promote uniform law.

364. E.C. Database Directive, *supra* note 17, at arts. 8(1), 15.

365. *Id.*, arts. 7(2), 7(5), 8(1).

366. P. Bernt Hugenholtz, *The New Database Right: Early Case Law from Europe*, paper presented at the Ninth Annual Conference on International Intellectual Property Law and Policy, Fordham University School of Law, New York (Apr. 19-20, 2001), available at <http://www.inir.nl/medewerkers/hughholtz.htm>; see also Reichman & Samuelson, *supra* note 18, at 87-95.

367. E.C. Database Directive, *supra* note 17, at art. 10(1).

provision of updates, can re-qualify that database *as a whole* for a new term of protection.³⁶⁸ In this and other respects, the scope of the *sui generis* adaptation right exceeds that of copyright law, which attaches only to the new matter added to an underlying, pre-existing work and expires at a time certain.³⁶⁹

Finally, the Directive carries no national treatment requirement into its *sui generis* component. Foreign database producers become eligible only if their countries of origin provide a similar form of protection or if they set up operations within the European Union.³⁷⁰ Non-qualifying foreign producers, however, may nonetheless seek protection for their databases under residual domestic copyright and unfair competition laws, where available.³⁷¹

The E.C.'s Directive on the Legal Protection of Databases thus broke radically with the historical limits of intellectual property protection in at least three ways. First, it overtly and expressly confers an exclusive property right on the fruits of investment as such, without predicating the grant of protection on any predetermined level of creative contribution to the public domain. Next, it confers this new exclusive property right on aggregates of information as such, which had heretofore been considered as unprotectible raw material or basic inputs available to creators operating under all other pre-existing intellectual property rights. Finally, it potentially confers the new exclusive property right in perpetuity, with no concomitant requirement that the public ultimately acquire ownership of the object of protection at the end of a specified period.³⁷² The Directive thus effectively abolishes the very concept of a public domain that had historically justified the grant of temporary exclusive rights in intangible creations.³⁷³

368. *Id.* at art. 10(3).

369. See 17 U.S.C. §§ 103, 302 (2000); Reichman & Samuelson, *supra* note 18, at 84-90.

370. E.C. Database Directive, *supra* note 17, at art. 11, Recital 56.

371. *Id.* at art. 13. Copyright protection is independent of *sui generis* protection, but the two regimes may apply cumulatively.

372. In this and other respects, the E.C. Database Directive broke with the history of intellectual property law by allowing a property rule—as distinct from a liability rule—to last in perpetuity. Trademarks may last in perpetuity, but they do not protect innovation or investments as such, only the signs and symbols that enable consumers to distinguish one producer's goods from another's. William Landes & Richard A. Posner, *Trademark Law: An Economic Perspective*, 30 J.L. ECON. 265 (1987). Trademarks are thus not legal monopolies, and because they protect only against acts that yield a likelihood of confusion, there are historical questions about their status as "property" at all.

These historical debates in turn reflect confusion about the fundamental distinction between exclusive property rights and liability rules, which have a different economic logic. See discussion *supra* at note 151. Trademarks are property in the sense that proprietors obtain legally enforceable entitlements, but that entitlement is only to avoid deceiving or confusing consumers by the adoption of similar identifying symbols. While the property-like status of marks has been strengthened against dilution in recent years, a trademark confers no rights in the underlying products of innovation or investment as such, which anyone remains free to copy and sell under a different mark.

373. Reichman & Samuelson, *supra* note 18. See generally James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 LAW & CONTEMP. PROBS. 33 (Winter/Spring 2003) [hereinafter Boyle, *Second Enclosure Movement*]; David Lange, *Recognizing the Public Domain*, 44 LAW & CONTEMP. PROBS. 147 (Autumn 1981); Jessica Litman, *The Public Domain*, 39 EMORY L.J. 965 (1990). But see van Caenegem, *supra* note 113, at 324, 328-30.

ii. The Database Protection Controversy in the United States

The situation in the United States differs markedly from that which preceded the adoption of the European Commission's Directive on the Legal Protection of Databases. In general, the legislative process in the United States has become relatively transparent. Since the first legislative proposal, modeled on the E.C. Directive, was introduced by the House Committee on the Judiciary in May of 1996,³⁷⁴ this transparency has generated a spirited and often high-level public debate.³⁷⁵ Very little progress toward a compromise solution had been reached as of the time of writing, however, which is hardly surprising given the intensity of the opposing views, the methodological distance that divides them, and the political clout of the opposing camps.³⁷⁶

We are accordingly left with the two basic proposals that were still on the table at the end of the last legislative session, which ended in an impasse. These proposals, as refined during that session, represent the baseline positions that each coalition carried into the current round of negotiations. One bill, H.R. 354, as revised in January of 2000,³⁷⁷ embodies the proponents' last set of proposals for a *sui generis* regime built on an exclusive property rights model (although some effort has been made to conceal that solution behind a facade that evokes unfair competition law). The other bill, H.R. 1858, sets out the opponents' views of a so-called minimalist misappropriation regime as it stood on the eve of the current round of negotiations.³⁷⁸

(a) *The exclusive rights model.* The proposals embodied in H.R. 354 attempt to achieve levels of protection comparable to those of the E.C. Directive by means that are more congenial to the legal traditions of the United States.³⁷⁹ The changes introduced in that bill softened some of the most controversial provisions at the margins, while maintaining the overall integrity of a strongly protectionist regime.³⁸⁰

The bill in this form continued to define "collections of information" very

374. H.R. 3531, 104th Cong. (1996). For details, see Reichman & Samuelson, *supra* note 18, at 102-09.

375. See J. H. Reichman, *Database Protection in a Global Economy*, *supra* note 18 at 467-70. For developments in the period between 1996 and 1999, see Reichman & Uhler, *Database Protection*, *supra* note 4, at 821-28.

376. Nevertheless, at the beginning of the last series of negotiations between the stakeholder groups in March 2001, the two committee chairmen vowed to draft a compromise bill if the interested parties themselves failed to agree. See Transcript of Press Conference of Rep. Billy Tauzin and Rep. James Sensenbrenner, March 29, 2001, available at <http://www.techlawjournal.com/cong107/database/20010329>.

377. See H.R. 354, 106th Cong. (1999). This bill was subject to proposed amendments on Jan. 11, 2000, which, however, were not formally submitted as an amended proposal. The summary in the text sometimes reflects changes that were introduced in publicly disclosed proposals for amendments.

378. See H.R. 1858, 106th Cong. (1999).

379. See generally Amanda Perkins, *United States Still No Closer to Database Legislation*, 2000 E.I.P.R. 366; Band & Kono, *supra* note 290; Roger L. Zissu, *Protection for Facts and Databases in the New World Order*, 1998 J. COPYRIGHT SOC'Y U.S.A. 271 (1998).

380. Many of these changes came under pressure from agents of the past Administration seeking to engender a compromise.

broadly as “information . . . collected and . . . organized for the purpose of bringing discrete items of information together in one place or through one source so that persons may access them.”³⁸¹ Like the E.C. Directive, the bill then casts eligibility in terms of an “investment of substantial monetary or other resources” in the gathering, organizing or maintaining of a “collection of information.”³⁸² It confers two exclusive rights on the investor: first, a right to make all or a substantial part of a protected collection “available to others;” and, second, a right “to extract all or a substantial part to make available to others.” Here the term “others” is manifestly broader than “public” in ways that remain to be clarified.³⁸³

H.R. 354 then superimposed an additional criterion of liability on both exclusive rights that is not present in the E.C. Directive. This is the requirement that, to trigger liability for infringement, any unauthorized act of “making available to others” or “extraction” for that purpose must cause “material harm to the market” of the qualifying investor “for a product or service that incorporates that collection of information and is offered or intended to be offered in commerce.” The crux of liability under the bill thus derives from a “material harm to markets” test that is meant to cloud the copyright-like nature of the bill³⁸⁴ and shroud it in different terminology.³⁸⁵

Here a number of concessions were made to the opponents’ concerns in the last public iteration of the bill on Jan. 11, 2000, some of them real, others nominal in effect. The addition of “material” to the market harm test,³⁸⁶ may, for example, address complaints that proponents viewed one lost sale as constituting actionable harm to the market.

At the same time, the revised bill contained convoluted and tortuous definitions of “market” that the previous Administration hoped would reduce the scope of protection in the case of follow-on applications.³⁸⁷ On closer inspec-

381. H.R. 354, 106th Cong. §1401(1) (1999). Here the overlap with copyright law is so palpable that it is hard to conceive of any copyrightable assemblage of words, numbers, facts, or information that would not also qualify as a potentially protectible collection of information.

382. *Id.* at §1402(a).

383. However, the second right represents a concession to the past Administration in that it foregoes the general *right to control private use* that appeared in previous versions. This concession thus *reduces* the scope of protection to a point more in line with the E.C.’s reutilization right, and it does not impede personal use by one who lawfully acquires access to the database. *See id.* at §1402(a).

384. Because the Supreme Court insists that “originality” is a constitutional requirement for copyright protection, it “has arguably stripped Congress of power . . . to enact a copyright-like statute for non-original databases.” Ginsburg, *U.S. Initiatives, supra* note 18, at 73-74.

385. *See* H.R. 354, 106th Cong. § 1402(a) (1999). As originally deposited, H.R. 354 spoke of “material harm to the primary market or a related market” of the investor. *Id.* The analysis in the text is based on the more refined but unpublished proposals of January 11, 2000. In fact, a “harm to markets” test is lifted bodily from section 107(4) of the Copyright Act of 1976, and it reflects the better view of what U.S. copyright law is all about. *See* J. H. Reichman, *Goldstein on Copyright Law: A Realist’s Approach to a Technological Age*, 43 STAN. L. REV. 943 (1991) (reviewing PAUL GOLDSTEIN, COPYRIGHT: PRINCIPLES, LAW AND PRACTICE (1990)).

386. H.R. 354 § 1402(a).

387. “Market” is thus supposed to assimilate “all markets” in which a protected investor “derives or reasonably expects to derive substantial revenue, directly or indirectly,” as well as all markets in which that investor “has taken demonstrable steps discernable to the public, to offer in commerce within a

tion, however, these definitions provide a static picture of a moving target that amounts to a mostly illusory limitation on the investor's broad adaptation right.³⁸⁸ Notwithstanding these so-called concessions, the bill effectively assigns most follow-on applications to any initial investor whose dynamic operations expand the range of potentially protectible matter with every update, *ad infinitum*.

The bill then introduced a "reasonable use" exception that was intended to benefit the nonprofit user communities, especially researchers and educators,³⁸⁹ and that conveys a sense of similarity to the fair use exception in copyright law.³⁹⁰ Once again, these benefits become largely illusory on closer analysis, because under the proposed bill, the very facts, data, and information that copyright law excludes have themselves become the objects of protection, and there are no other significant exceptions. Hence, virtually every customary or traditional use of facts or data compiled by others that copyright law would presumably have allowed scientists, researchers, or other nonprofit entities to make in the past now becomes a *prima facie* instance of infringement under H.R. 354. These users would, in effect, either have to license such uses or be prepared to seek judicial relief for "reasonableness" on a continuing basis. Because university administrators dislike litigation and are risk averse by nature, and this provision puts the burden of showing reasonableness on them, there is reason to expect a chilling effect on customary uses by these institutions of data heretofore in the public domain.³⁹¹

short period of time a product or service" from which he expected to derive substantial revenue. H.R. 354 §§ 1401(3)(A), (B) (with additional proviso added Jan. 11, 2000).

388. In principle, only actual, likely, or planned markets are protected under this scheme, which creates a narrow opening for a value-adding competitor who arrives on the scene with an unlikely or unplanned application. Even here, however, the definitions ignore the prospects that the initial investor will continue to expand the range of projected investments over time and thus convert all the tests to moving targets that constantly expand his potential claims to protected market segments.

In practice, moreover, database proprietors would be well-advised to plan for any market segments they can remotely foresee over time and to craft their business models in broadly worded terms accordingly. Should a competitor discover a surprise market niche to slip into all the same, the initial proprietor's most likely strategy would be to surround the competitor with applications of its own, to limit the competitor's field of expansion and to extract cross-licenses wherever possible.

389. H.R. 354 § 1403(2).

390. In copyright law, there is a thicket of exclusions and exceptions that must be worked through before anyone can infringe. In particular, one cannot infringe for a taking of unprotectible facts or ideas, and even a taking of protectible expression may be excused by codified exceptions for, say, teaching or research. The "fair use" exception comes into play only as a last resort, to excuse marginal takings by an alleged infringer that advance the public interest at a small cost to the proprietor. *See* 17 U.S.C. §§ 102(b), 107-122 (2000).

391. *Cf.* Ginsburg, *U.S. Initiatives*, *supra* note 18, at 69; Reichman & Uhler, *Database Protection*, *supra* note 4, at 812-20, 825-29. A further provision then completes the sense of circularity by expressly exempting any nonprofit educational, scientific, and research use that "does not materially harm the market" as previously defined. *See* H.R. 354 § 1403(b). Since any use that does not materially harm the market remains unactionable to begin with, this "concession" adds nothing but window dressing. However, another vaguely worded exception seems to recognize at least a possibility that certain "fully transformative uses" might nonetheless escape liability, but this ambiguous proposal defies interpretation in its present form and remains to be clarified.

The bill recognized an “independent creation” norm, which presumably exempts any database, however similar to an existing database, that was not the fruit of “copying.”³⁹² This provision codifies a fundamental norm of copyright law, and the European Commission made much of a similar norm in justifying its own regulatory scheme. In reality, this “independent creation” principle produces unintended and socially deleterious consequences when transposed to the database milieu precisely because many of the most complex and important databases are inherently incapable of independent regeneration. Sometimes the database cannot be reconstituted because the underlying phenomena are one-time events, as often occurs in the observational sciences.³⁹³ In other instances, key components of a complex database can no longer be reconstituted with certainty at a later date. Any independently regenerated database suffering from these defects would necessarily contain gaps that made it inherently less reliable than its predecessor.

These problems point to a more general phenomenon that affects competition in large or complex databases. Even when, in principle, such databases could be reconstituted from scratch, the high cost of doing so—as compared with the add-on costs of existing producers—will tend to make the second-comer’s costs so high as to constitute a barrier to entry. Meanwhile, the first-comer’s comparative advantage from already owning a large collection that is too costly to reconstitute will only grow more formidable over time, an economic reality that progressively strengthens the barriers to entry and tends to reinforce (and, indeed, to explain) the predominance of sole-source data suppliers in the marketplace.³⁹⁴

Government-generated data remained excluded, in principle, from protection, in keeping with current U.S. practice,³⁹⁵ which differs from E.U. practice in this important respect. However, there is considerable controversy surrounding the degree of protection to be afforded government-generated data that subsequently become embodied in value-adding, privately funded databases.³⁹⁶ All parties agree that a private, value-adding compiler should obtain whatever degree of protection is elsewhere provided, notwithstanding the incorporation of government-generated data, assuming that this transaction entails a “substantial investment.”³⁹⁷ The issue concerns the rights and abilities of third parties to continue to access the original, government-generated data sets. The propo-

392. H.R. 354 § 1403(c).

393. See Reichman & Uhler, *Database Protection*, *supra* note 4, at 807-08.

394. See Benkler, *Free as the Air*, *supra* note 91. As more and more segments of industry come to appreciate the market power that major database producers could acquire under the proposed legislation, one after another has petitioned the subcommittee for special relief. Thus, H.R. 354, which grew to some thirty pages in length, singled out various special interests that benefit, to varying degrees, from special exemptions from liability. At the time of writing the list of those entitled to such immunities included news reporting organizations; churches that depend on genealogical information, notably the Mormons; online service providers; and certain online stockbrokers. See H.R. 354 § 1403(e)(f)(i).

395. See H.R. 354 § 1404.

396. See, e.g., Weiss & Backlund, *supra* note 7, at 300, 303.

397. For the view that it may not, see Ginsburg, *U.S. Initiatives*, *supra* note 18, at 70.

nents of H.R. 354 have been little inclined to accept measures seeking to preserve access to the original data sets, despite pressures in this direction.³⁹⁸

H.R. 354 imposed no restrictions whatsoever on licensing agreements, including agreements that might overrule the few exceptions otherwise allowed by the bill.³⁹⁹ Despite constant remonstrations from opponents about the need to regulate licensing in a variety of circumstances—and especially with respect to sole-source providers⁴⁰⁰—the bill itself does not budge in this direction. On the contrary, new provisions added to H.R. 354 in 2000 would set up measures that prohibit tampering with encryption devices (“anti-circumvention measures”) and electronically embedded “watermarks” in a manner that parallels the provisions adopted for online transmissions of copyrighted works under the DMCA.⁴⁰¹ Because these provisions would effectively secure the database against unauthorized access (and so tend to create an additional “exclusive right of access” without expressly so declaring),⁴⁰² they would only add to the database owner’s market power to dictate contractual terms and conditions without regard to the public interest. These powers are further magnified by the imposition of criminal sanctions in addition to strong civil remedies for infringement.⁴⁰³

The one major concession that was made to the opponents’ constitutional arguments concerns the question of duration. As previously noted, the E.C. Directive allows for perpetual protection of the whole database so long as any part of it is updated or maintained by virtue of a new and substantial investment, and the proponents’ early proposals in the United States echoed this provision.⁴⁰⁴ However, the U.S. Constitution clearly prescribes some limited term of duration for intellectual property rights,⁴⁰⁵ and the proponents have finally bowed to pressures from many directions by limiting the term of duration to fifteen years.⁴⁰⁶

Any update to an existing database would then qualify for a new term of fifteen years, but this protection would apply, at least in principle, only to the matter added in the update. In practice, however, the inability to clearly separate old from new matter in complex databases, coupled with ambiguous language concerning the scope of protection against harm to “likely, expected,

398. See A QUESTION OF BALANCE, *supra* note 1, at 102-05.

399. H.R. 354 § 1404(e).

400. For a recent discussion, see Ginsburg, *U.S. Initiatives*, *supra* note 18, at 76.

401. Cf. Digital Millennium Copyright Act (DMCA), Pub. L. No. 105-304, § 103, 112 Stat. 2860, 2863 (1998) (codified at 17 U.S.C. §§ 512, 1201, 1205). For a discussion of the DMCA, see *supra* notes 286-287 and accompanying text.

402. See, e.g., Ginsburg, *U.S. Initiatives*, *supra* note 18, at 63-68; Jane C. Ginsburg, *Copyright and Control Over New Technologies of Dissemination*, 101 COLUM. L. REV. 1613 (2001).

403. See H.R. 354 §§ 1406-1407.

404. See H.R. 3534 and H.R. 2652; Reichman & Samuelson, *supra* note 18, at 103-09 (citing authorities).

405. U.S. Const., art. I, § 8, cl. 8; *Eldred v. Reno*, 239 F.3d 372 (D.C. Cir. 2001), *cert. granted sub nom.*, *Eldred v. Ashcroft*, 534 U.S. 1126, *and cert. amended*, 534 U.S. 1160 (2002).

406. See H.R. 354 § 1409(i).

or planned” market segments,⁴⁰⁷ may still leave a loophole for an indefinite term of duration.⁴⁰⁸

(b) *The unfair competition model.* The opponents’ bill, the Consumer and Investor Access to Information Act of 1999, H.R. 1858, was introduced by the House Commerce Committee in 1999, as a sign of good faith,⁴⁰⁹ in response to critics’ claims that the opponents’ coalition sought only to block the adoption of any database protection law.⁴¹⁰ H.R. 1858 begins with a definition of databases that is not appreciably narrower than that of H.R. 354, except for an express exclusion of traditional literary works that “tell a story, communicate a message,” and the like.⁴¹¹ In other words, it attempts to draw a clearer line of demarcation between the proposed database regime and copyright law, to reduce overlap or cumulative protection as might occur under H.R. 354.

The operative protective language in H.R. 1858 appeared short and direct, but it relied on a series of contingent definitions that muddy the true scope of protection. Thus, the bill would prohibit anyone from selling or distributing to the public a database that is (1) “a duplicate of another database . . . collected and organized by another person or entity,” and (2) “is sold or distributed in commerce in competition with that other database.”⁴¹² The bill then defines a prohibited duplicate as a database that is “substantially the same as such other database, as a result of the extraction of information from such other database.”⁴¹³

Here, in other words, liability attached only for a wholesale duplication of a pre-existing database that resulted in a substantially identical end product. However, this basic misappropriation approach becomes further subject to both expansionist and limiting thrusts. Expanding the potential for liability is a proviso added to the definition of a protectible database that treats “any discrete sections [of a protected database] containing a large number of discrete items of information” as a separably identifiable database entitled to protection in its own right.⁴¹⁴ The bill would thus codify a surprisingly broad prohibition of

407. *Id.*

408. *Cf.* Ginsburg, *U.S. Initiatives*, *supra* note 18, at 76 (proposing measures to close this loophole).

409. *See* H.R. 1858, 106th Cong. (1999); *see also* H.R. REP. NO. 106-350, PART I (1999); Perkins, *supra* note 379.

410. This may be true of some, but not all, members of that coalition. Both scientists and universities, for example, although allied with the opponents’ coalition for strategic reasons, prefer a minimalist approach because they want some protection against unauthorized commercial applications of their data without hindering access to data for public research activities. Over time, pressures for some form of database protection have built up to the point where the minimalist alternative bill has become a serious basis of negotiation, even though, in our view, it remains poorly crafted and contains numerous ambiguities.

411. *See* H.R. 1858 § 101(1).

412. *See id.* § 102.

413. *See id.* § 101(2).

414. *See id.* § 101(1)(B).

follow-on applications that make use of discrete segments of pre-existing databases,⁴¹⁵ subject to the limitations set out below.

A second protectionist thrust resulted from the lack of any duration clause whatsoever, with the prohibition against wholesale duplication—subject to limitations set out below—*conceivably lasting forever*. This perpetual threat of liability would attach to wholesale duplication of even a discrete segment of a pre-existing database, if the other criteria for liability were met.

These powerfully protective provisions, put into H.R. 1858 at an early stage to weaken support for H.R. 354, were offset to some degree by other express limitations on liability and by a codified set of misuse standards to help regulate licensing. To understand these further limitations, one should recall that liability even for wholesale duplication of all, or a discrete segment, of a protected database does not attach unless the unauthorized copy is sold or distributed in commerce and “in competition with” the protected database.⁴¹⁶ The term “in competition with,” when used in connection with a sale or distribution to the public, is then defined to mean that the unauthorized duplication “displaces substantial sales or licenses likely to accrue from the original database” and “significantly threatens . . . [the first-comer’s] opportunity to recover a reasonable return on the investment” in the duplicated database.⁴¹⁷ Both prongs must be met before liability will attach.

It follows that even a wholesale duplication that was not commercially exploited or did not substantially decrease expected revenues (as might occur from, for example, nonprofit scientific research activities) could presumably escape liability in appropriate circumstances. Similarly, a follow-on commercial product that made use of data from a protected database might escape liability if it were sold in a distant market segment or required substantial independent investment.

H.R. 1858 then further reduced the potential scope of liability by imposing a set of well-defined exceptions and limiting enforcement to actions brought by the Federal Trade Commission (“FTC”).⁴¹⁸ There are express exceptions comparable to those under H.R. 354 for news reporting, law enforcement activities,

415. The Clinton Administration expressed reservations about this de facto derivative work right, which is built into a regime that lasts forever. Communication with Professor Justin Hughes (Oct. 17, 2001) (on file with authors). At that time, Professor Hughes worked at the U.S. Patent and Trademark Office and was the principal coordinator of the Administration’s submission of opinions regarding the database protection bills introduced in Congress.

416. See H.R. 1858 § 102.

417. *Id.* § 101(5).

418. See *id.* § 107. The provision that conditions liability for infringement on an official FTC action was a tactical expedient devised to provide the House Commerce Committee with some basis for asserting concurrent jurisdiction over database legislation, along with that of the House Judiciary Committee’s Subcommittee on Courts and Intellectual Property. Most observers believe that the absence of any private right of action in H.R. 1858 as it stands constitutes a fatal flaw that would have to be removed in any final compromise decision to adopt an unfair competition approach. Some supporters consider FTC supervision a necessary safeguard, especially in view of the First Amendment tensions that any database protection law is certain to generate in the United States.

intelligence agencies, online stockbrokers, and online service providers.⁴¹⁹ There is also an express exception for nonprofit scientific, educational, or research activities,⁴²⁰ in case any such uses were thought to escape other definitions that limit liability to unauthorized uses in competition with the first-comer. Still other provisions clarify that the protection of government-generated data or of legal materials in value-adding embodiments remains contingent upon arrangements that facilitate continued public access to the original data sets or materials.⁴²¹ A blanket exclusion of protection for “any individual idea, fact, procedure, system, method of operation, concept, principle or discovery” wisely attempts to provide a line of demarcation with patent law and to ward off unintended protectionist consequences in this direction.⁴²²

Another important set of safeguards emerged from the drafters’ real concerns about potential misuses of even this so-called “minimalist” form of protection. These concerns are expressed in a provision that expressly denies liability in any case where the protected party “misuses the protection” that H.R. 1858 affords. A related provision then elaborates a detailed list of standards that courts could use as guidelines to determine whether an instance of misuse had occurred.⁴²³ These guidelines or standards would greatly clarify the line between acceptable and unacceptable licensing conditions, and if enacted, they could make a major contribution to the doctrine of misuse as applied to the licensing of other intellectual property rights as well.⁴²⁴

In summary, the underlying purpose of H.R. 1858 was to prohibit wholesale duplication of a database as a form of unfair competition. It thus set out to create a minimalist liability rule that prohibits market-destructive conduct rather than to enact an exclusive property right as such,⁴²⁵ and in this sense, initially posed a strong contrast to H.R. 354. Over time, however, different iterations of the bill, designed to win supporters away from H.R. 354, have made H.R. 1858 surprisingly protectionist—especially in view of its de facto derivative work right.⁴²⁶

419. H.R. 1858 §§ 103(b), 103(c), 104(b), 104(e), 106(a).

420. *See id.* § 103(d).

421. *See id.* §§ 101(b), 104(f). There are also express exclusions of telecommunications carriers’ subscription lists (for example, telephone directories) and of securities market data. *Id.* § 104(g). However, the bill proposes an amendment to the Securities Exchange Act of 1934 that would prohibit the misappropriation of “real-time” stock market information. *Id.* § 201.

422. *See id.* § 104(d).

423. *Id.* §§ 106(b), 106(b)(1-6). *See infra* note 540.

424. These provisions could particularly assist judicial regulation of shrink-wrap and click-on licenses affecting online distribution of software and other electronic information tools. *See generally* Reichman & Franklin, *supra* note 16, at 929-60.

425. For early proposals to this effect, see Reichman & Samuelson, *supra* note 18, at 139-45. For a definition of liability rules, see *supra* note 151.

426. The realities of the bargaining process are such that concessions made to the high-protectionist camp at an earlier stage, for whatever tactical reasons, are unlikely to be withdrawn later.

C. Implications for Science: Disintegration of the Research Commons?

Part II of this study described some of the potentially limitless possibilities for research and innovation that might ensue from using digital technologies to exploit scientific data available from the public domain as it was traditionally constituted. However, these prospects dim the moment we consider the ramifications for science of the economic, legal, and technological assaults on the public domain currently under way. This section explores some of the likely negative effects that these trends could have on science and innovation unless science policy directly addresses these risks.

1. Restricting Access to and Use of Scientific Data

In the interests of clarity, we outline the implications of present trends on a sectoral basis, in keeping with the functional map of public domain data flows indicated above. We begin with the government's role as primary producer of such data and then consider the implications of present trends for academia and the private sector.

a. In Government

If a basic trend is the shifting of more data production and dissemination activities from government to the private sector, one should recognize at the outset that the social benefits of such a shift can exceed its costs under the right set of circumstances. In principle, private database producers may sometimes operate more efficiently and attain qualitatively better results than government agencies. Positive effects are especially likely when markets have formed, competition occurs, and the public interest, including the needs of the research community that was previously served by the government, continues to be satisfied.

There are also numerous drawbacks associated with this trend, however, which require careful consideration.⁴²⁷ To begin with, a private data supplier will seldom be in a position to produce the same quantity and range of data as a government agency, charge prices that users can afford, and still make a profit. In other words, a government agency has typically taken on the task of data production and dissemination precisely because the social need for such data outweighs the market opportunities for these activities. Social costs from privatization begin to rise if the profit motive induces a private supplier to unduly reduce the quantity and range of data produced or made available.

For example, a private data producer typically markets refined data products to end users in relatively small quantities, whereas basic research, particularly in the observational sciences, often requires raw or less commercially

427. See, e.g., DAVID, DIGITAL BOOMERANG, *supra* note 20, at 2 (“We really do not know how much further the current rush toward privatization of scientific and technological knowledge can go before it starts to seriously undermine the inherited structure of fragile conventions and institutions that support cooperative research activities, thereby setting in motion the contraction of the global domain of scientific inquiry.”).

refined data in voluminous quantities. On the whole, overzealous privatization of the government's data production capabilities poses real risks for both science and innovation because the private sector simply cannot duplicate the government's public good functions and still make a profit.⁴²⁸

Unless the private sector can demonstrably produce and distribute much the same data more effectively and with higher quality standards than a government agency, privatization may become little more than a sham transaction. The would-be entrepreneur merely appropriates a government function and then licenses data back to a captive market at much higher prices and with greatly increased restrictions on access and use. In the absence of market-induced competition, there is a very high risk of trading one monopolist with favorable policies toward science and the broader society—the government—for another monopolist driven entirely by profit and the restrictions made necessary by this motive.

Absent a sham transaction, one cannot say *a priori* that any given privatization project necessarily results in a net social loss. The outcome will depend on the contracts the agency stipulates and the steps it is willing to take to ensure continued access to data for research purposes on reasonable terms and conditions. In contrast to buying data collection services, the licensing of data and information products from the private sector raises serious questions about the type of controls the private sector places on the redistribution and uses of such data and information the government can subsequently undertake. If the terms of the license are onerous to the government, and access, use, and redistribution are substantially restricted—as they almost always are—neither the agency nor the taxpayer is well served. This is particularly true in cases where the data that need to be collected are for a basic research function or serve a key statutory mission of the agency.

A classic example of what can go wrong was the privatization of the Landsat earth remote sensing program in the mid-1980s. Following the legislatively mandated transfer of this program to EOSAT Co., the price per scene rose more than 1000%, and significant restrictions were imposed, even on research uses. Use by both government and academic scientists fell sharply, and recent studies have shown the extent to which both basic and applied research in environmental remote sensing was set back.⁴²⁹ This experiment also failed in com-

428. See Stephen Maurer, *Raw Knowledge: Protecting Technical Databases for Science and Industry*, in NATIONAL RESEARCH COUNCIL, COMMITTEE FOR A STUDY ON PROMOTING ACCESS TO SCIENTIFIC AND TECHNICAL DATA FOR THE PUBLIC INTEREST, PROCEEDINGS OF THE WORKSHOP ON PROMOTING ACCESS TO SCIENTIFIC AND TECHNICAL DATA FOR THE PUBLIC INTEREST: AN ASSESSMENT OF POLICY OPTIONS (Jan. 14-15, 1999), available at http://www.nap.edu/html/proceedings_sci_tech/appC.html (last visited Feb. 14, 2003). (“Since public databases do not generate the revenues needed to pay license fees, statutory protection imposes substantial (albeit inadvertent) pressures to privatize.”).

429. See BITS OF POWER, *supra* note 1, at 121-24.

mercial terms, as EOSAT Co. became unable to continue operations after several years.⁴³⁰

Several lessons can be drawn from this and similar undertakings. One is that before a transfer to the private sector occurs, objective criteria demonstrating net social gains from the transaction should be met. Recent studies have identified such criteria, and when they are met, good reasons to privatize would exist.⁴³¹ Even when objective criteria justify a transfer of data production from the public to the private sector, government agencies should not abdicate their contractual responsibility to ensure access for research purposes on favorable terms and conditions, as will be discussed further in the final Part of this article.

We concede that, if the government lacks resources to generate data for a public research function in the first place, obtaining it from a willing producer is better than nothing.⁴³² Where, however, a choice exists, the wrong decision can impose high opportunity costs on the scientific community and the broader society. We also assume the government intends to continue its policy of not commercializing its own data output—unlike the European Union and most other countries—in view of the positive externalities this has generated in the past.

b. In Academia

The legal and technological pressures identified above will also affect the uses that are made of government-funded data in academic and other nonprofit institutions. These pressures will intensify the tensions that already exist between the sharing norms of science and the need to restrict access to data in pursuit of increased commercial opportunities.

Although the enhanced opportunities for commercial exploitation that new intellectual property rights and related developments make possible are clear, they will affect the normative behavior of the scientific community gradually and unevenly. Academics are already conflicted in this emerging new environment, and these conflicts are likely to grow.⁴³³ As researchers, they need continued access to a scientific commons on acceptable terms, and they are expected to contribute to it in return. As members of academic institutions, however, they are increasingly under pressure to transfer research results to the private sector for gain, and they themselves may want to profit from the new commercial opportunities.⁴³⁴

430. The government then reassumed control of this program, but a new privatizing proposal is under consideration. See Landsat Data Continuity Mission, at <http://ldcm.usgs.gov/> (last updated Jan. 24, 2003).

431. See BITS OF POWER, *supra* note 1, at 116-24; RESOLVING CONFLICTS, *supra* note 8, at 82-87.

432. *But see* Stephen Maurer & Suzanne Scotchmer, *Database Protection: Is It Broken and Should We Fix It?*, SCIENCE, May 14, 1999, at 1129 (noting that funding agencies are notoriously reluctant to pay for intellectual property licenses).

433. See generally Rai, *supra* note 33, at 109-15 (describing normative changes in academia after Bayh-Dole and countervailing efforts to redress the balance).

434. See Powell, *supra* note 13.

The government itself fuels these conflicts through the potentially contradictory policies that underlie its funding of research. One message reminds scientists of their duties to share and disclose data, in keeping with the traditional norms of science. The other, more recent message urges them to transfer the fruits of their research to the private sector or otherwise exploit the intellectual property protection their research may attract.⁴³⁵

At the moment, these conflicts are strongest where the line between basic and applied science has collapsed and commercial opportunities inhere in most projects.⁴³⁶ Obvious examples are biotechnology and computer science. There, progress frequently occurs through accretions of know-how, obtained by trial and error, and theoretical explanations may follow, rather than precede, practical applications.⁴³⁷ Decisions about the use of intellectual property rights and licensing contracts to exploit applications thus rebound in unexpected ways against the possibilities of further research.

In the future, the enactment of a powerful intellectual property right in collections of data might be expected to push these tensions into other areas where the lines between basic and applied research remain somewhat clearer and the pressures to commercialize research results have been less noticeable thus far. In exploring the implications of these developments for academic research, we continue to focus attention on the two distinct but overlapping research domains previously characterized as “formal” and “informal.”

i. The Formal Zone

In what we call the formal sector, science is conducted within structured research programs that establish guidelines for the production and dissemination of data. Typically, data are released to the public in connection with the publication of research results.⁴³⁸ Data may also be disclosed in connection with patent applications and supporting documentation. One should recall that, even without regard to the mounting legal and technological pressures, there are strong economic pressures that already limit the amount of data investigators are inclined to release at publication or in patent applications. There are growing delays in releasing those data as researchers consider commercialization options, and more of the data that are released come freighted with various restrictions.⁴³⁹

435. See, e.g., Rai, *supra* note 33, at 112 (noting that the initial impetus to commercialize and corresponding “norm change has . . . been tempered by some attention to the traditional values of research science”).

436. See Eisenberg, *Public Research*, *supra* note 120, at 1726 (“[A]s university patenting and private funding of university research increase, the time-honored distinction between ‘basic’ and ‘applied’ research is becoming ever more difficult to maintain, particularly in fields that are of significant commercial interest.”).

437. See, e.g., Reichman, *Computer Programs*, *supra* note 292. See generally Reichman, *Legal Hybrids*, *supra* note 164.

438. See *supra* note 72 and accompanying text.

439. *Id.*; see also David Blumenthal et al., *Withholding Research Results in Academic Life Sciences: Evidence from a National Survey of Faculty*, 92 JAMA 1224 (1997); Rai, *supra* note 33, at 110-11.

The enactment of a hybrid intellectual property right in collections of data, such as the E.C. Directive, would introduce a disruptive new element into an already troubled academic environment. Suddenly, such a right would make it possible to publish academic research for credit and reputation while retaining ownership and control of the underlying data, which would no longer automatically lapse into the public domain. By the same token, disclosure of research results for the purpose of filing patent applications, while continuing to count as novelty-defeating prior art in the public domain,⁴⁴⁰ would not displace the inventor's right to control the underlying data that support the application. Because patent law has been encroaching progressively on collections of data that scientists previously regarded as falling within the public domain,⁴⁴¹ the database right itself—depending on how it was structured—could become more valuable than patent protection.

To some extent, this development tends to erase some of the preexisting distinctions between the “formal” and “informal” domains. In both domains, access to data might increasingly have to be secured by means of brokered, negotiated transactions, and this outcome is rife with implications.⁴⁴² For present purposes, it seems clear that any database protection law, coupled with the other legal and technological measures discussed above, will further undermine the sharing ethos and encourage the formation of a strategic, self-interested trading mentality that already predominates in the informal domain.

These pressures will necessarily tend to blur and dilute the importance of publication as the line of demarcation between a period of exclusive use in relative secrecy and ultimate dedication of data to the public. Once databases attract an exclusive property right valid against the world, the legal duty of scientists publishing research results to disclose or release the underlying data could depend on codified exceptions permitting use for verification and certain “reasonable” nonprofit research and educational purposes.⁴⁴³ Of course, this new default rule would ultimately have to be reconciled in practice with the disclosure obligations of the federal funding agencies. The point is that the new default rule would nonetheless place even published data outside of the public domain, and much academic research is not federally funded or funded in ways that waive such disclosure requirements.

The role of academic journal publishers in this new legal environment also warrants consideration. At present, scientists tend to assign their copyrights to publishers on an exclusive basis, and many of these journals now produce electronic versions—sometimes exclusive of a print version. These practices

440. 35 U.S.C. § 102 (2000).

441. See, e.g., Eisenberg, *Public Research*, *supra* note 120 (stressing the changing interpretation of boundaries of intellectual property law); Eisenberg, *Bargaining*, *supra* note 13; Rai, *supra* note 33.

442. “Scientists report having to wait months or even years to carry out experiments, while their institutions attempt to negotiate the terms of ‘Material Transfer Agreements,’ . . . database access agreements, and patent license agreements.” Eisenberg, *Bargaining*, *supra* note 13, at 225.

443. See *supra* notes 334-36 and accompanying text. We assume some such provision would necessarily be incorporated into any U.S. database law that Congress may adopt.

already complicate matters because, as shown above, the data that traditional copyright law puts into the public domain may be fenced to a still unknown extent by the technological measures that the DMCA reinforces. If, in addition, a database law is enacted, any data that the scientist assigns to the publisher with the article would become subject to the new statutory regime. The publisher would then be in a position to control subsequent uses of the data and to make them available online under a subscription or pay-per-use plan with additional restrictions on extraction or reuse.

Even if individual scientists are willing and able to resist the demands for exclusive assignment of both their copyrights and any new database rights, the fact remains that publication of the article in a journal would no longer automatically release the data into the public domain. On the contrary, and unless the scientist waived the new default rule, even the data revealed in the publication itself would remain subject to his or her exclusive right of extraction and reuse—at least as formulated under the E.C. database protection directive.⁴⁴⁴

With or without a new statutory database right in the United States, scientists appear certain to come under increasing pressure to retain data for commercial exploitation. The research universities are already deeply committed to maximizing income from patentable inventions under Bayh-Dole, with varying degrees of success, and they will logically extend these practices and procedures to the commercialization of databases as valuable research tools.⁴⁴⁵ A key question is whether they will make the commercialized data available for academic research on reasonable terms and conditions.⁴⁴⁶

As with government-generated data, university efforts to commercially exploit their databases could produce net social gains under the right set of circumstances. Besides the incentive to generate new and more refined data products that an intellectual property right might confer, greater efforts could be made to enhance the quality and utility of selected databases than might otherwise be the case. Absent such incentives, many scientists may not take pains to organize and document their data for easy use by others, particularly those outside their immediate disciplines, and may not refine their data beyond the level needed to support their own research needs and related publication objectives. Legal incentives might thus stimulate the production of more refined databases, especially where markets for such products had formed.

At the same time, these new commercial opportunities would tempt university administrators and academics to attenuate or modify the sharing and open access norms of science and to circumvent obligations in this regard that federal

444. These exclusive rights would then remain subject to any exceptions for scientific use incorporated into the database protection right. These exceptions were negligible in the E.C. Database Directive, *supra* note 17, and efforts to obtain significantly broader exceptions in the U.S. proposals have not succeeded so far. See Reichman & Uhlir, *Database Protection*, *supra* note 4, at 825-27.

445. This process is already underway, even in the absence of any database protection right. See, e.g., Powell, *supra* note 13, at 254-55, 263-65; Eisenberg, *Bargaining*, *supra* note 13, at 224-26.

446. “University technology transfer professionals report that agreements presented for the transfer of research tools impose increasingly onerous terms.” Eisenberg, *Bargaining*, *supra* note 13, at 225.

agencies had established.⁴⁴⁷ Were this to occur, the unintended harm to research could greatly exceed that we are accustomed to experiencing with regard to patented inventions under Bayh-Dole⁴⁴⁸ because the licensing of academic databases, reinforced by a codified intellectual property right, would limit the quantity and quality of data heretofore available from the public domain.

From a qualitative perspective in particular, the data produced at universities has typically been more refined or highly processed than government data and are developed with particular research objectives or applications in mind.⁴⁴⁹ Moreover, many of these data-intensive research activities require access to and use of multiple sources of data.⁴⁵⁰ What has been changing is the evident commercial value of this type of refined, upstream research byproduct that makes databases both outputs as well as inputs at a much earlier stage of the research process in many areas of science. This raises serious doubts about their continued availability on acceptable terms or whether they will even be made available to other researchers at all.

Present university licensing practices with regard to material transfer agreements (“MTAs”) in the biotechnology sector do not bode well in this regard. These contracts are not drafted by scientists with the needs of the larger scientific community in mind. Recent surveys of these practices show that university technology licensing offices resort to exclusive licenses that impose onerous terms and conditions, including aggressive grant-back or reach-through clauses that attempt to secure a share of the return from follow-on applications developed with the aid of the licensed technologies.⁴⁵¹ While anecdotal and some empirical research suggests that these offices have showed a certain willingness to negotiate reasonable terms in specific instances, the legal and economic literature foresees growing anticommons effects and ever-higher transac-

447. Cf. Rai, *supra* note 33, at 109-11 (stressing that commercial involvement in academic research has already “undermined norms governing the sharing of research materials and tools”).

448. See, e.g., Eisenberg, *Public Research*, *supra* note 120, at 1667 (“[B]y providing incentives to patent and restricting access to discoveries made in institutions that have traditionally been the principal performers of basic research, [Bayh-Dole] threatens to impoverish the public domain of research science that has long been an important resource for researchers in both the public and private sectors.”).

449. See Nelson, *supra* note 132, at 16 (“By far the lion’s share of modern scientific research, including research done at universities, is in fields where a practical application is central in the definitions of a field. In today’s world science is useful to inventing, not so much because of Serendipity, but because much of modern scientific research is designed exactly to help clear the path for technological progress.”).

450. See *Collections Hearings*, *supra* note 183.

451. See, e.g., Eisenberg, *Bargaining*, *supra* note 13, at 225 (stressing onerous terms and limitations and overwhelming “burden of reviewing and renegotiating each of a rapidly growing number of agreements for what used to be routine exchanges among scientists.”); Rai, *supra* note 33, at 111 (noting that many MTAs require researchers to assign or license intellectual property rights to discoveries made in the course of using the research tools, while others “prohibit sharing tools with other researchers or sending them to other institutions”).

tion costs.⁴⁵² There are also enormous opportunity costs for research that will prove difficult or impossible to document.

There is no reason to expect that, left to their own devices, the university technology licensing offices dealing on a case-by-case basis would demonstrate any greater concern for the research needs of the larger community with respect to databases than they have with respect to MTAs. In fact, many university databases could become more valuable than the corresponding patent portfolios over time, owing to their cumulative nature, the potential ability to control updates under the proposed new intellectual property right, and the existing ability to control online dissemination and use by electronic contracts. As the need to exploit such databases upstream becomes more pronounced, with corresponding palpable commercial payoffs, university administrators could logically become less willing to make commercially valuable data available, even to colleagues, in the absence of corresponding benefits.

If these predictions prove even partly accurate, we should then expect to see the formation of university “database pools” and cross-licensing agreements, like the “patent pools” of today,⁴⁵³ which can achieve some positive synergies through cooperation.⁴⁵⁴ However, the evidence shows that such pools are very difficult to form when the value of upstream research products defies easy measurement and the relevant players in a given industry have very different agendas, as would occur when federal agencies, academic institutions, and different types of private companies are all involved.⁴⁵⁵ Moreover, there are far

452. See, e.g., Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, SCIENCE, May 1, 1998, at 698. But see J. Walsh, et al., *Research Tool Patenting and Licensing and Biomedical Innovation*, in W. COHEN & S. MERRILL, PATENTS AND THE KNOWLEDGE BASED ECONOMY (forthcoming 2003) (finding no hard evidence of anticommons effects in downstream patents, but emerging problems affecting research in upstream patents).

453. Describing Patent pools, Professor Rai says:

Patent pools typically function by extending membership to those firms in a given industry that agree to assign or license their individual patent rights to the pool. In simple pools, membership gives each party the right to royalty-free licensing of all patents in the pool. In more sophisticated pools, members who use a particular patent pay the pool a set fee that reflects the economic significance of the patent; similarly royalties accumulated by the pool are divided according to the perceived significance of the technology put in by the parties.

Rai, *supra* note 33, at 129.

454. See, e.g., Eisenberg, *Bargaining*, *supra* note 13, at 227-28 (“Collaborative research that pools research capabilities and funds from different institutions in the public and private sectors is increasingly common, not only in the life sciences but across all fields of research.”); Gregory Graff & David Zilberman, *Towards an Intellectual Property Clearinghouse for Ag-Biotechnology*, 3-2001 IP STRATEGY TODAY 1, 9 (2001) (“[A]n intellectual property ‘clearinghouse’ might be a most effective way to reduce market inefficiencies that hinder the exchange of privately deemed knowledge, allowing researchers to obtain the freedom-to-operate status necessary to commercialize . . . research.”); Merges, *Patent Pools*, *supra* note 151, at 123, 155-56. See generally Robert P. Merges, *Contracting Into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CAL. L. REV. 1293 (1996) [hereinafter Merges, *Contracting into Liability Rules*].

455. Rai, *supra* note 33, at 133. Professor Rai, focusing on the example of prospective biotech patent pools, points out that the relevant academic institutions and federal agencies might logically want to make marketable end products “widely available nonexclusively on a low or no-royalty basis,” while “the private companies focused exclusively on upstream research might believe in more selective licensing at a higher royalty.” *Id.*

greater risks that such pools lead to collusive, anti-competitive behavior, to the erection of formidable barriers to entry, and to discrimination, which in this case could adversely affect lower-tier universities that possessed few tradable assets.⁴⁵⁶

At present, the primary bulwarks against such a breakdown of the sharing ethos are the formal requirements of the federal funding agencies, which in many cases continue to require that data from the research projects they fund be transferred at some point to public repositories or made available upon request. To avoid these negative results, the agencies would have to strengthen these requirements—and their enforcement—and adapt them to the emerging high-protectionist intellectual property environment. We elaborate further on this topic in Part IV. The point for now is that, absent express overrides that universities voluntarily adopt or that funding agencies impose in their research grants and contracts, the new default rules of ownership and control would automatically take effect if Congress enacts a database protection law, and they could become general practice even without such a law as the result of routine, unregulated database licensing practices.

These new default rules of ownership and control could gradually undermine and dissolve the pre-existing norm that scientists publish and release their data to the public. No well-meaning resolution to the contrary by scientific bodies will, by itself, avert this outcome.

ii. The Informal Zone

In the informal zone, researchers are not yet ready to publish or are working independently on small science projects beyond the formal controls and requirements of a federal research program mandating open access or public deposit. If the research project is federally funded, the investigator is still operating in the pre-publication phase, in which he enjoys a period of exclusive use that typically lasts from six months to two years, depending on the grant. Any formal obligations to disclose data that derive from the grant do not yet apply, and data exchanges in this phase depend on self-interest, competitive advantages, and the sharing ethos. In addition, much of the research falling within the informal zone is funded by state governments, foundations, and the universities themselves, all of which leave more discretion in these matters to researchers, as well as by private companies, who normally require secrecy.

Our concerns about the effects of the new legal and technological pressures on the formal academic zone apply with even greater force to the informal zone where the impetus to commercialize data will encounter fewer regulatory constraints. The changing mores likely to undermine disclosure and open access in the formal zone would make it harder to organize cooperative networks in the

456. See, e.g., Rai, *supra* note 33, at 130-35 (stressing ability of pools to withhold access to those that lack important technology to contribute or that will not pay large licensing fees). See generally H. HOVENKAMP, M.D. JANIS, & M.A. LEMLEY, IP AND ANTITRUST §§ 34.3, 34.4 (2002).

less structured and more unruly informal domain.⁴⁵⁷

This loss of cooperative incentives would prove troublesome even if the informal zone were to remain stable in the face of these pressures. In reality, such pressures—and especially a new intellectual property right in databases—will seriously destabilize the informal zone as depicted in recent sociological studies. As the new default rules make themselves felt, researchers operating in the informal zone will become aware that they own property interests in their data collections over and above any stipulated obligation to publish. In other words, a self-conscious assertion of property rights—and the corresponding proprietary mentality—will displace the softer, more inchoate legal norms that otherwise protect confidential information in ways that scientists in the informal domain typically know little about.

At present, researchers in the informal zone tend to accommodate requests to share data in response to community norms, peer pressure, the expectation of reciprocity, and other factors shaped by perceived self-interest. Under a strong database protection regime, researchers will logically begin to view such transactions as requests to waive or relinquish exclusive property rights, whose potential value is not easily measured or foreseen.⁴⁵⁸ This outlook would tend to make researchers more reluctant to dilute their rights and more inclined to hoard data, demand up-front short term benefits as a quid pro quo, and even to insist on their own versions of the reach-through and grant-back clauses that are already routinely used by university technology licensing offices. As university administrators become increasingly aware of the commercial possibilities inherent in database protection, they may restrict their academics' freedom to informally exchange data with colleagues and require university approval—lest such exchanges damage their potential commercial interests.⁴⁵⁹ Moreover, private partners that support university research will insist that statutory property rights in data be fully respected, and more and more private partners may become involved in the commercialization of data produced by academics.⁴⁶⁰

At the very least, the informal data exchanges of the past, which were already hampered by various forms of personal strategic considerations, seem likely to become more formal and complicated, with higher transaction costs and real risks of encountering holdouts, thickets of overlapping rights, and anti-

457. See Hilgartner, *Access to Data*, *supra* note 31, at 9 (noting that IP protection will make collaboration more difficult); Rebecca S. Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. CHI. L. REV. 1017, 1061 (1989); Rai, *supra* note 33, at 129-35 (documenting difficulties of organizing collective exchange norms to address transaction costs).

458. On the difficulties of evaluation, see Rai, *supra* note 33, at 125-27.

459. See, e.g., Powell, *supra* note 13, at 264-65 (stressing the extent to which universities are “seeking to privatize new information” and fearing erection of “costly toll booths”).

460. Cf. Rai, *supra* note 33, at 182-83 (finding that since Bayh-Dole and related legal developments in the 1980s and 1990s, industry now funds “a non-trivial percentage of academic research in the life sciences,” some of these relationships “resemble commercial joint ventures,” and participants “often depart quite markedly from traditional research norms”).

commons effects.⁴⁶¹ To be sure, the advent of an exclusive property right in non-copyrightable databases might facilitate some transactions that could not have occurred in the past, owing to legal uncertainty.⁴⁶² It may also stimulate new types of transactions by reinforcing the trading mentality and encouraging parties to seek deals based on their data assets, although such incentives and potential benefits are far better suited to private sector activities than to the academic milieu. On the whole, however, the outcome is likely to be increased obstacles to the construction of informal academic networks of data exchanges, with a corresponding reduction in the flow of the data streams discussed in Part II.⁴⁶³ Individual researchers will be strongly tempted to hold out or bargain to impasse, at the expense of scientific cooperation.⁴⁶⁴

Examples of such phenomena are already observable. In academia, one promising initiative to organize a common database of human mutations data on a quasi-commercial basis, while maintaining broad community access, failed for a variety of anticommons effects, and the contributing entities ultimately bargained to impasse.⁴⁶⁵

If the residual force of the sharing ethos in the informal sector started to break down under these pressures, the process of disintegration would encounter fewer bulwarks protecting the public domain than in the formal sector. Scientists operating in the informal zone are, by definition, less constrained by formal federal data access requirements, and they are often closer to industry. Indeed, the more the cooperative spirit dissipates, the more likely it becomes that the commercial ethos of the private sector will fill the vacuum and pervade the informal domain.

These tendencies would predictably become more pronounced over time, as more scientists become aware of the new possibilities to retain ownership and control of data, even after publication of research results. Indeed, one would logically expect that strategic behavior in the informal zone would increasingly be geared to efforts to maximize advantages from post-publication opportunities. Should this occur, academics themselves would exert pressure on the fed-

461. See, e.g., Heller & Eisenberg, *supra* note 452; Powell, *supra* note 13, at 264; cf. Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, 1 INNOVATION POL. & ECON. 119 (2001).

462. But see Mowery & Rosenberg, *supra* note 231, at 49 (finding that, in “biotechnology, continuing uncertainty over the strength and breadth of intellectual property protection may have discouraged litigation”).

463. See, e.g., Powell, *supra* note 13, at 266 (fearing “turf wars, with rival networks of partners looking to delay, deter, and defend themselves against competitors and poachers rather than advancing both their efforts and those of the overall field”).

464. See, e.g., *id.* at 265 (stressing risk of “disputes, duplication, and discord”); cf. Rai, *supra* note 33, at 127-29 (stressing risk of holdups in biotechnology because patentee seeks to appropriate much of the value of the improvements).

465. See Stephen M. Maurer, *Inside the Anticommons: Academic Scientists’ Struggle to Commercialize Human Mutations Data, 1999-2001*, paper given at the Franco-American Conference on the Economics, Law, and History of Intellectual Property Rights, Haas School of Business, University of California at Berkeley, Oct. 5-6, 2001 [hereinafter Maurer, *Inside the Anticommons*].

eral system and their universities to fall in line with the needs of commercial partners.

One can thus project a kind of cascading effect if a strong database protection right were enacted and the scientific community failed to take steps to preserve and reinforce the research commons. On this view, today's formal zone, built around the release of data into the public domain at publication, would begin to resemble the informal zone, as sociologists have recently portrayed it,⁴⁶⁶ while that same informal zone would look more and more like the private sector. Under these circumstances, one cannot necessarily assume the open access policies currently supporting the formal sector would continue in force, in which case even basic research could be adversely affected—as occurred in the United Kingdom in the 1980s through 1990s.⁴⁶⁷

What the new equilibrium—resulting from the conflict between these privatizing and commercializing pressures on the one hand and the traditional norms of public science on the other—will look like cannot be predicted with any degree of certainty. In a previous article, however, we outlined the cumulative negative effects such tendencies would likely have on scientific endeavor. For the sake of brevity, they are recalled here in summary form:

- (1) Less effective domestic and international scientific collaboration, with serious impediments to the use, reuse, and transformation of factual data that are the building blocks of research;
- (2) Increased transaction costs driven by the need to enforce the new legal restrictions on data obtained from different sources, and the implementation of new administrative guidelines concerning institutional acquisitions and uses of databases, and associated legal fees;
- (3) Monopoly pricing of data and anti-competitive practices by entities that acquire market power, or by first-entrants into niche markets that predominate in many research areas; and
- (4) Less data intensive research and opportunity costs.⁴⁶⁸

What could well be the greatest casualty of all are the new opportunities that digital networks provide to create virtual information commons within and across discipline-specific communities built around optimal access to and exchange of scientific data. To the extent that public science becomes dominated by brokered intellectual property transactions, the resulting combination of high transaction costs, unbridled self-interest, and anticommons effects could defeat the fragile cooperative arrangements needed to create and maintain such

466. See generally Hilgartner & Brandt-Rauf, *Controlling Data*, *supra* note 31.

467. See, e.g., P.J. Runei, *Energy R&D in the United Kingdom*, BATTELLE, Mar. 2000, at 7 (paper prepared for the U.S. Department of Energy under contract DE-AC06-76RLO 1830) (citing government statistics showing that the U.K. government has downsized its support for R&D over the past two decades and that the British government has acknowledged that this policy has not worked well and that the “country’s R&D had reached a critical level of ill health and a condition that threatens future economic growth and international competitiveness”).

468. Reichman & Uhlir, *Database Protection*, *supra* note 4, at 812-21.

virtual information commons and the distributed research opportunities they make possible.

In industry. Proponents of a strong database protection law claim it is needed to stimulate investment in more databases than would otherwise become available.⁴⁶⁹ However, there is no credible evidence that the market for databases has been under-supplied or under-invested in the United States, even though the share of U.S. commercial databases in the world market has declined somewhat in the last ten years.⁴⁷⁰ On the contrary, the European Union's production since the enactment of the Directive, which reportedly showed an initial short-term spike, has subsequently remained stable.⁴⁷¹

The emergence of digitally networked environments "has generated a host of new value-added services and products, and appreciably increased the importance of this segment of the database market."⁴⁷² In a previous article, Professors Reichman and Samuelson explained why digital technology would cause the market for value-added database products to flourish in the near future,⁴⁷³ and their predictions have held up over time. The database industry as a whole, and its value-adding components, have in fact flourished, despite constant allegations of market failure.⁴⁷⁴

It remains, of course, logical to consider whether that industry's long-term growth prospects would suffer in the absence of additional legal protection against free-riding expropriations from databases that were costly to develop and maintain. In so doing, however, one must discount the availability and effectiveness of self-help measures,⁴⁷⁵ and the relative social costs of removing vast quantities of technological information from the public domain, which have functioned as basic inputs of the knowledge economy.⁴⁷⁶ At the very least, these and other considerations should focus attention on the choice of legal instru-

469. See, e.g., E.C. Database Directive, *supra* note 17, Recitals 6-12; Maurer, *Across Two Worlds*, *supra* note 34 (critically evaluating this thesis); *Information Antipiracy Hearings*, *supra* note 239; BRAUNSTEIN, *supra* note 239; see also Hunsucker, *supra* note 239.

470. Maurer, *Across Two Worlds*, *supra* note 34, at 11.

471. *Id.* at 35-40.

472. Fewer, *supra* note 341, at 177.

473. First, "digital technologies facilitate the disaggregation of value-added functions" and permit new forms of data aggregation and presentation that were unavailable in print media. Reichman & Samuelson, *supra* note 18, at 125. Second, "digital technologies foster new functions, such as reformatting, filtering, and hot-linking, which have no counterparts in print media." *Id.*

474. See Fewer, *supra* note 341 (finding no evidence of market failure in Canada); Maurer, *Across Two Worlds*, *supra* note 34, at 11. See generally A QUESTION OF BALANCE, *supra* note 1.

475. For example, electronic fencing through encryption devices, coupled with tagging or watermarking of data, make it possible for online database providers to impose standardized contractual restrictions on all would-be users. See A QUESTION OF BALANCE, *supra* note 1, at 68-69; Kenneth W. Dam, *Self-Help in the Digital Jungle*, in EXPANDING THE BOUNDARIES OF IP, *supra* note 13, at 103, 107-10, 117-19.

476. See, e.g., Boyle, *Second Enclosure Movement*, *supra* note 373.

ments to remedy any perceived market failure, and on the relative social costs and benefits of different approaches.⁴⁷⁷

For present purposes, it suffices to emphasize that any de facto exclusive property right in the non-copyrighable contents of databases would automatically empty the public domain of most of the factual matter the *Feist*⁴⁷⁸ decision had consigned to that domain in 1991. While government-generated data would probably remain available to the public under domestic (but not foreign) law,⁴⁷⁹ all non-governmental databases would become presumptively proprietary, whether made available online or in hard copies. Access to, and use of, such data for research purposes would depend on negotiated licenses or on any research exceptions to the proprietary rights that happened to be adopted in the end.⁴⁸⁰ Depending on the type and strength of the database protection law ultimately adopted, there are good reasons to fear that barriers to entry could be high, unlicensed follow-on applications could be stifled, and sole-source providers would likely predominate.⁴⁸¹

While industry would thus contribute significantly less aggregate data to the public domain than in the past,⁴⁸² its ability to bank on these same proprietary rights might induce the private sector to disclose certain kinds of data previously kept secret, especially to potential partners in the universities. Much would depend on the willingness and ability of the academic community to accommodate the private sector's needs to restrict access to or use of data made available for nonprofit research pursuits⁴⁸³ and to deny access or use to would-be competitors.

To the extent that a more protectionist database regime would facilitate more public-private partnerships with universities, the social benefits likely to ensue from such interaction would have to be weighed against the risks that the norms of industry would increasingly pervade academia and foster pressure for less public disclosure of data, in potential conflict with the norms of science. At the very least, industry usually insists on confidentiality in joint projects, and would further want any end product to benefit from any new database protec-

477. Critics argue that, given the power of self-help remedies in the digital environment, contract and unfair competition law would suffice to close any regulatory gaps that were likely to ensue in the short or medium term, without further encumbering access to the public domain. See e.g., Bott, *supra* note 34; Reichman & Samuelson, *supra* note 18, at 137-50.

478. *Feist Publ'ns, Inc. v. Rural Tel. Servs. Co.*, 449 U.S. 340 (1991).

479. All previous versions of database protection legislation introduced in Congress have exempted both federal and state collections of information from the scope of protection.

480. See Reichman & Uhler, *Database Protection*, *supra* note 4, at 822-28; *supra* notes 362-363, 389-391, 420, and accompanying text.

481. See generally Reichman & Uhler, *Database Protection*, *supra* note 4, at 813-21.

482. This is clear under the E.C. Database Directive, where duration is potentially limitless in time. See E.C. Database Directive, *supra* note 17. Where duration is limited, long-term gains of public domain data are at least conceivable, at the expense of short and medium-term losses. See, e.g., van Caenegem, *supra* note 113, at 325.

483. Cf. Rai, *supra* note 33, at 111 ("Companies that license . . . [biotech] materials and tools to academic researchers often force researchers to sign material transfer agreements (MTAs) that tightly restrict the researchers' use of these materials.").

tion right that Congress ultimately enacted.⁴⁸⁴ To parry these and other anticipated negative developments, concerted efforts to accommodate and protect the research goals of government funding agencies, university administrations, and academic scientists would have to be made. Detailed proposals to this effect are set out in Part IV.

Major repercussions from database protection would also be felt in those sectors of the economy where subpatentable innovation depends on the constant exchange of technical information and know-how among the members of engineering communities working on given technical trajectories.⁴⁸⁵ As shown above, this vibrant component of the innovation economy currently depends on liability rules that protect confidential information and on a robust public domain in which members of the technical community exchange sub-patentable know-how and information that automatically enters the public domain with disclosure or independent creation.⁴⁸⁶ The advent of a strong exclusive database right that displaced the existing pro-competitive liability rules could hamper these exchanges, reduce spillover effects, hinder value-adding innovation, and elevate the costs of R&D, all of which could slow the pace of sub-patentable innovation generally.

2. Broader Implications for the Innovation System

Much of the economic literature that so far has addressed the topic of database protection tends unconsciously to assume the premises that ultimately yield the authors' expected conclusions. Because most economists uncritically equate "property rights" with "exclusive rights," and because the risk of market failure inherent in public goods is often efficiently overcome with "property rights," these studies usually end where they began: by endorsing property rights and taking the view that stronger is better.⁴⁸⁷ Such studies beg the important questions that a deeper knowledge of intellectual property law might raise, namely, what level and mode of protection would produce the greatest amount of investment with the most acceptable degree of social costs.⁴⁸⁸

a. Underestimating the Potential Social Costs

In this connection, a codified, federal unfair competition law, based on the misappropriation rationale, could constitute a minimalist response to a poten-

484. *Id.* at 111 (noticing that corporate sponsors of research demand secrecy); *cf.* Nelson, *supra* note 132, at 33 ("Discussions with industry executives suggest that, until recently, industry often gave research a *de facto* research exemption. However, now they often are very reluctant to do so. In many cases they see university researchers as direct competitors to their own research efforts aimed to achieve a practical result which is patentable. And they feel themselves burdened by the requirement to take out licenses to use university research results that are patented, and see no reason why they shouldn't make the same demands on universities.").

485. *Cf.* Reichman, *Of Green Tulips*, *supra* note 164; Samuelson & Scotchmer, *supra* note 155.

486. *See supra* notes 163-167 and accompanying text.

487. *See, e.g., Information Antipiracy Hearings*, *supra* note 239; BRAUNSTEIN, *supra* note 239. For trenchant criticism of this approach, see Boyle, *Cruel, Mean or Lavish*, *supra* note 165.

488. *See, e.g.,* Rochelle C. Dreyfuss, *Information Products: A Challenge to Intellectual Property Theory*, 20 N.Y.U. J. INT'L L. & POL. 897 (1988).

tial gap in the law whose true dimensions remain unknown. It could also provide the uniform model needed for proper administration of the national system of innovation and for negotiating an international arrangement.⁴⁸⁹ Moreover, a growing number of innovative proposals rooted in liability rules⁴⁹⁰ have been put on the table in recent years, in addition to the better-known proposals for a more traditional unfair competition approach.⁴⁹¹

Most economists engaged in this topic, however, have so far ignored these and other proposals largely because their economic models and premises either do not allow them to take liability rules into account or incline them to postulate their inherent inferiority to exclusive rights.⁴⁹² In so doing, they fail to devote any serious attention to the social costs that critics of such rights—including strong database protection—continue to fear. As a result, formal economic analysis has so far taught us very little about how to craft an alternative protective regime so as to avoid market failure without erecting barriers to entry and impoverishing the public domain.

The most fundamental question these economists largely ignore is the extent to which any exclusive property right might *a priori* constitute the wrong kind of solution for a legal regime that aims to protect investment in large-scale aggregates of data as such.⁴⁹³ Consider, for example, the interactions that might occur once the line between patentable inventions and the new intellectual property right in data began to blur. In the past, patent disclosures entered the public

489. See, e.g., Reichman, *Database Protection in a Global Economy*, *supra* note 18, at 493-96 (discussing a model international treaty based on the repression of wholesale duplication of databases under a menu of legal options); Zissu, *supra* note 379.

490. See definition of liability rules *supra* note 151.

491. For example, Wendy Gordon has proposed a tort of “malcompetitive copying” that would rest on specific economic criteria. Gordon, *On Owning Information*, *supra* note 152; see also Karjala, *Misappropriation*, *supra* note 152. William Kingston has proposed a new type of liability regime that would transform intellectual property protection from a duration-based calculus of rights to an accounting-based calculus of rights based on multiples of R&D costs. William Kingston, *Unlocking the Potential of Intellectual Property*, paper presented to Swedish International Symposium on Economics, Law and Intellectual Property, Gotheberg 4 (June 26-30, 2000); see also WILLIAM KINGSTON, *THE DIRECT PROTECTION OF INNOVATION* 1-124 (William Kingston ed., 1987). J. H. Reichman and Pamela Samuelson have elsewhere proposed a “compensatory liability” regime that would allow second-comers freely to extract data from a protected database to develop value-adding follow-on products, so long as adequate compensation was paid under an “automatic license” (not a compulsory license) for a specified period of time. See Reichman & Samuelson, *supra* note 18, at 145-51; see also Reichman, *Of Green Tulips*, *supra* note 164.

492. See, e.g., Robert P. Merges, *Of Property Rules, Coase, and Intellectual Property*, 94 COLUM. L. REV. 2655 (1994); Merges, *Contracting Into Liability Rules*, *supra* note 454. On this view, even the new transactional economic theories, which do worry about the use of intellectual property rights after they are granted, are miraculously advanced by a total dependence on exclusive rights, despite the need for patent pools and other institutions of dubious social value, to attenuate anticommons and other anti-competitive effects. See, e.g., Merges, *Patent Pools*, *supra* note 151, at 124-32.

493. Cf. Yochai Benkler, *A Political Economy of the Public Domain: Markets in Information Goods Versus the Marketplace of Ideas*, in EXPANDING THE BOUNDARIES OF IP, *supra* note 13, at 267, 269-74. One would, indeed, expect or prefer economic analysis to focus on the comparative advantages and disadvantages of using either exclusive property rights or liability rules to address the underlying risks of market failure. See, e.g., Calabresi & Melamed, *supra* note 151.

sector immediately after the patent issued.⁴⁹⁴ Patented inventions expressed in claims approved by the patent office expired after twenty years,⁴⁹⁵ and any relevant sub-patentable know-how remained subject to reverse engineering by “proper means” under the liability rules that protect trade secrets.⁴⁹⁶

If a strong database right were eventually enacted, however, the ability of a second-comer to practice the claims to a patented invention that nominally expired could in fact depend on his or her gaining access to underlying data in a database that the patent holder, or his assignees, continued to hold under the database right. Even if the original data supporting the claims also lapsed under the database right,⁴⁹⁷ the patent holder could attempt to generate new data to surround the original patent claims in order to make the second-comer’s exercise of follow-on improvements more difficult.⁴⁹⁸

If the initial patent claims were narrowly drawn, in keeping with present day approaches to many biotech, software, and business method patents,⁴⁹⁹ the patentee’s independent rights in data might enable him or her to project an aura surrounding these narrow claims that would actually magnify the exclusionary power of the resulting patents. Conversely, when broad patent claims were allowed, as has allegedly occurred with respect to some biotech research tools,⁵⁰⁰ there are already well-known risks of impeding follow-on applications and blocking progress.⁵⁰¹ An exclusive property right in data could further magnify these risks. It would expose second-comers to further allegations of unlawfully using data from databases that surround and integrate these claims and thus reinforce the social disutilities already associated with broad patent claims.

Moreover, even as regards sub-patentable innovation, once the aggregates of information that constitute an entrepreneur’s technical know-how were reduced to data embodied in a protectible database, the resulting proprietary

494. If the patent failed to issue, the traditional U.S. rule preserved trade secrecy. 35 U.S.C. § 122(a) (2000). However, the virtually universal practice is to publish the contents of patent applications after eighteen months, and U.S. law has begun to conform to this practice in recent years. *See* 35 U.S.C. §§ 122(b)(1)(A), 122(b)(2).

495. 35 U.S.C. §§ 112, 154(2).

496. UNIF. TRADE SECRETS ACT (UTSA), §1(4), comment, 14 U.L.A. 449 (1985) (“Proper means include . . . [d]iscovery by ‘reverse engineering’ . . .”).

497. As must occur in the United States, but not in the European Union.

498. In this event, the database right would generate some of the problems currently associated with so-called “blocking patents.” *See, e.g.,* Robert Merges, *Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents*, 62 TENN. L. REV. 75 (1994) [hereinafter Merges, *Blocking Patents*].

499. *See, e.g.,* *Diamond v. Diehr*, 450 U.S. 175 (1981); *Diamond v. Chakrabarty*, 447 U.S. 303 (1980); *State St. Bank & Trust Co. v. Signature Fin. Group*, 149 F.3d 1368 (Fed. Cir. 1998).

500. *See, e.g.,* Rai, *supra* note 33, at 104 (stating that “[m]any . . . [expressed sequence tag (“EST”)] applications are notable for the broad scope of their patent claims: the applications claim not only the EST but also the full genre of which it is a part and future uses of the gene”) (citing Christopher Anderson, *A New Model for Gene Patents*, 260 SCIENCE 23 (1993)). Some companies have reportedly filed patent applications on hundreds or thousands of ESTs. Rai, *supra*, at 104.

501. *See generally* Eisenberg, *Bargaining*, *supra* note 13, at 225-26, 231-47; Heller & Eisenberg, *supra* note 452; Rai, *supra* note 33, at 126; Lemley, *Improvement*, *supra* note 288, at 1053-54; Merges, *Blocking Patents*, *supra* note 498.

rights could make it much harder or more costly for third parties to obtain that know-how by reverse engineering.⁵⁰² Similarly, the database right could be used to discourage efforts to work around patents or to add value to either patented inventions or sub-patentable innovations.⁵⁰³ Over time, if the cumulative database rights extended into related fields of innovation, they could become the “wheel” that actually governed the “spokes,” and in this sense, acquire more value—and impose more anti-competitive effects—than a given firm’s patent portfolio.

One could thus conceive of an interlocking web of data rights that enabled a proprietor with strategic patents and copyrights to surround and control a spectrum of knowledge in given fields. On this scenario, the database proprietors could agglomerate the prior art into a kind of expanding arctic shelf of privatized information, with ever lower costs for aggregators, ever higher costs for users in the absence of competition, and ever higher barriers to entry. In such a case, antitrust laws might provide the only form of relief, and that is always a cumbersome and uncertain course of action.⁵⁰⁴

It is not that these negative synergies are certain to occur, but rather that these potential social costs must be weighed against any social benefits thought to derive from a strong exclusive property right in collections of data. Because the social costs of striking the wrong balance are manifestly so high, and the uncertainties attendant upon database protection are so great, the most credible economic advice has been that of Maurer and Scotchmer, who advise against taking any premature action that might make the end result far worse than the predicament from which we started.⁵⁰⁵

b. A Market-Breaking Approach

While the E.U. authorities proclaim the success of their Directive,⁵⁰⁶ the evidence is inconclusive and at most supports a finding that the Directive has, as yet, failed to produce the harmful long-term consequences that critics expect.⁵⁰⁷ The list of critics who predict such consequences has grown, however, and the longer the *sui generis* database law is implemented in practice, the more likely its socially harmful, over-protectionist consequences will become evident.

502. Cf. Samuelson & Scotchmer, *supra* note 155.

503. Cf. Reichman, *Of Green Tulips*, *supra* note 164.

504. See, e.g., Arti K. Rai, *Fostering Cumulative Innovation in the Biopharmaceutical Industry: The Role of Patents and Antitrust*, 16 BERKELEY TECH. L.J. 813 (2001); Hanns Ullrich, *Intellectual Property, Access to Information, and Antitrust: Harmony, Disharmony, and International Harmonization*, in EXPANDING THE BOUNDARIES OF IP, *supra* note 13, 35, 381-98. “Too much or too easy reliance on antitrust relief will never solve deficiencies that the intellectual property system is likely to show.” *Id.* at 383.

505. Maurer & Scotchmer, *supra* note 432.

506. See Maurer, *Across Two Worlds*, *supra* note 34.

507. This section is based on Reichman, *Database Protection in a Global Economy*, *supra* note 18, at 482-84.

To see why critics in the United States and elsewhere⁵⁰⁸ harbor deep concerns about the long-term consequences of the E.U.'s approach, it suffices to grasp how radical a change it would introduce into the U.S. system of innovation and to consider how great the risks of such change really are. Traditionally, United States intellectual property law did not protect investment as such—a tradition that still has constitutional underpinnings.⁵⁰⁹ At the same time, the national system of innovation depends on enormous, upstream flows of mostly government-generated or government-funded scientific and technical information, which everyone is free to use,⁵¹⁰ and on free competition with respect to downstream information goods.

The domestic intellectual property laws protected downstream bundles of information in two situations only: copyrightable works of art and literature, and patentable inventions. However, the following conditions apply in both cases:

- (1) These regimes require palpably significant creative contributions based on free inputs of information and ideas.
- (2) They presuppose a flow of unprotected information and data upstream.
- (3) They presuppose free competition with regard to the products of mere investment that are neither copyrightable nor patentable.⁵¹¹

As previously observed, the E.C.'s Database Directive changes this approach, as would the parallel proposal, H.R. 354, to enact strong database rights in the United States. Specifically, these *sui generis* regimes confer a strong and, in the European Union, potentially perpetual exclusive property right on the fruits of mere investment, without requiring any creative contribution. They also convert data and information—the previously unprotectible raw materials and basic inputs of the modern information economy—into the subject matter of this new exclusive property right.

The *sui generis* database regimes would thus effectuate a radical change in the economic nature and role of intellectual property rights (“IPRs”). Until now, the economic function of IPRs was to make markets possible where previously there existed a risk of market failure due to the public good nature of intangible creations. Exclusive rights make embodiments of intangible public goods artificially appropriable, create markets for those embodiments, and make it possible to exchange payment for access to these creations.⁵¹²

508. For a recent discussion, see MARK DAVISON, LEGAL PROTECTION OF DATABASES (forthcoming 2003), which is highly critical of the E.C. Directive. Davison is Associate Professor of Law, Monash University, Australia.

509. See, e.g., *Wal-Mart Stores, Inc. v. Samara Bros.*, 529 U.S. 205 (2000); *Bonito Boats, Inc. v. Thunder Craft Boats Inc.*, 489 U.S. 141 (1989); *Compeco Corp. v. Day-Brite Lighting, Inc.*, 376 U.S. 234 (1964); *Sears, Roebuck & Co. v. Stiffel Co.*, 376 U.S. 225 (1964); *Kellogg Co. v. Nat'l Biscuit Co.*, 305 U.S. 111 (1938) (rejecting *Int'l News Serv. v. Assoc. Press*, 248 U.S. 215 (1918)).

510. See A QUESTION OF BALANCE, *supra* note 1, at 4-8, 9, 52-58; Mowery & Rosenberg, *supra* note 231, at 40-48, 55-61.

511. See Reichman & Uhlir, *Database Protection*, *supra* note 4, at 832-38.

512. See, e.g., COOTER & ULEN, *supra* note 212, at 108-09, 126-35.

In contrast, an exclusive IPR in the contents of databases *breaks* existing markets for downstream aggregates of information that were formed around inputs of information largely available from the public domain. It conditions the very existence of all traditional markets for intellectual goods on:

- (1) The willingness of information suppliers to supply at all (they can hold out or refuse to deal);
- (2) The willingness of suppliers not to charge excessive or monopoly prices (i.e., more than downstream aggregators can afford to pay in view of their own risk management assessment); and
- (3) The willingness and ability of suppliers to pool their respective chunks of information in contractually constructed cooperative ventures.

This last constraint is perhaps the most telling of all. In effect, the *sui generis* database regimes create new and potentially serious barriers to entry to all existing markets for intellectual goods owing to the multiplicity of new owners of upstream information in whom they invest exclusive rights—any one of whom can hold out and all of whom can impose onerous transaction costs (analogous to the problem of expressed sequence tags (“ESTs”) and single nucleotide polymorphisms (“SNPs”) in patent law).⁵¹³ This thicket of rights fosters anticommons effects,⁵¹⁴ and the database laws appear to be ideal generators of this phenomenon.

In short, under the new *sui generis* database regimes, there is a built-in risk that too many owners of information inputs will impose too many costs and conditions on all the information processes we now take for granted in the information economy. At best, the costs of R&D activities seem likely to rise across the entire economy, well in excess of benefits, owing to the potential stranglehold of data suppliers on raw materials. This stranglehold will increase with market power if many databases are owned by sole-source providers. Over time, the comparative advantage from owning a large, complex database will tend progressively to elevate barriers to entry.⁵¹⁵

The potential social gains of a strong database law cannot justify incurring these risks of disrupting or deforming the national system of innovation. It hardly seems logical to break up all existing markets for intellectual goods just to cure an alleged market failure for investments in a single type of intellectual good, that is non-copyrightable collections of information. At present, the United States dominates this market, and there is no credible empirical evidence of market failure that could not be cured by more traditional means.⁵¹⁶

513. See, e.g., Rai, *supra* note 33, at 125-28.

514. See Heller & Eisenberg, *supra* note 452.

515. Accord Benkler *Constitutional Bounds*, *supra* note 91. All nonprofit activities will be especially hard hit. Over time, lost opportunity costs in neglected R&D projects owing to these balkanized inputs could become staggering, and many forms of innovation may stagnate as a result. Even so, it will not be easy to document these lost opportunity costs, although the past experience of science in this regard will be repeated across the whole information economy. For details, see Reichman & Uhler, *Database Protection*, *supra* note 4, at 812-20.

516. See, e.g., Fewer, *supra* note 341; Maurer, *Across Two Worlds*, *supra* note 34.

The foregoing analysis reinforces the hypothesis that an exclusive property right is the wrong way to address the problem of legal protection for electronic databases, and it reconfirms the desirability of considering a modern liability rule that could avoid market failure without impoverishing the public domain.⁵¹⁷ Supporters of strong database protection laws (and strong contractual regimes to reinforce them) believe that the benefits of private property rights are without limit, and that more is always better.⁵¹⁸ They expect that these powerful legal incentives will attract huge resources to the production of electronic information tools.⁵¹⁹

In contrast, critics fear that an exclusive property right in non-copyrightable collections of data, coupled with the proprietors' unlimited power to impose electronic adhesion contracts in the course of online delivery, will compromise the operation of the national system of innovation, which depends on the free flow of upstream data and information. In place of the explosive production of new databases that proponents envision, opponents of a strong database right predict a steep rise in the cost of information across the global information economy and a progressive balkanization or feudalization of that economy,⁵²⁰ in which fewer knowledge goods may be produced as more tithes have to be paid to more and more information conglomerates along the way.⁵²¹ In the critics' view, the information economy most likely to emerge from an exclusive property right in data will resemble models already familiar from the Middle Ages, when goods flowing down the Rhine River or moving from Milan to Genoa were subject to dozens, if not hundreds, of gatekeepers demanding tribute.

IV

A CONTRACTUALLY RECONSTRUCTED RESEARCH COMMONS FOR SCIENCE AND INNOVATION

The foregoing exposition has described the growing efforts underway to privatize and commercialize scientific and technical information that was heretofore freely available from the public domain or on an open access basis. If these pressures continue unabated, they will result in the disruption of long-established scientific research practices and in the loss of new research opportunities that digital networks and related technologies make possible. We do not expect these negative synergies to occur all at once, however, but rather to manifest

517. For details, see Reichman & Samuelson, *supra* note 18, 145-51.

518. See, e.g., Raymond Nimmer, *Breaking Barriers: The Relation Between Contract and Intellectual Property Law*, 13 BERKELEY TECH. L.J. (1998); Maureen A. O'Rourke, *Property Rights and Competition on the Internet: In Search of an Appropriate Analogy*, 16 BERKELEY TECH. L. J. 561 (2001).

519. See E.C. Database Directive, *supra* note 17, Recitals 6-12; *Information Antipiracy Hearings*, *supra* note 239; Braunstein, *supra* note 239.

520. Cf. DRAHOS & BRAITHWAITE, *supra* note 266, at 1-3 (stressing the risks of raising the costs of borrowing ideas and information so high that they "will progressively choke innovation," and warning that "[m]ost businesses will be losers, not winners").

521. Accord Maurer, *Across Two Worlds*, *supra* note 34; Maurer & Scotchmer, *supra* note 432; Benkler, *Constitutional Bounds*, *supra* note 91.

themselves incrementally, and the opportunity costs they are certain to engender will be difficult to discern.

Particularly problematic is the uncertainty regarding the specific type of database protection that Congress may enact and any exceptions favoring scientific research and education that such a law might contain.⁵²² As we have tried to demonstrate, moreover, the economic pressures to privatize and commercialize upstream data resources will continue to grow in any event.⁵²³ Legal means of implementing these pressures already exist, regardless of the adoption of a *sui generis* database right.⁵²⁴ Therefore, given enough economic pressure, that which could be done to promote strategic gains will likely be done by some combination of legal and technical means.

If one accepts this premise, then the enactment of some future database law could make it easier to impose restrictions on access to and use of scientific data than at present, but the absence of a database law or the enactment of a lower protectionist version of it would not necessarily avoid the imposition of similar restrictions by other means. In such an environment, the existing elements of risk or threat to the sharing norms of public science can only increase unless the scientific community adopts countervailing measures.

We accordingly foresee a transitional period in which the negative trends identified above will challenge the cooperative traditions of science and the public institutions that have reinforced those traditions in the past, with uncertain results. In this period, a new equilibrium will emerge as the scientific community becomes progressively more conflicted about their private interests and their communal needs for data and technical information as a public resource. This transitional period will provide a window of opportunity that should be used to analyze the potential effects of a shrinking public domain and to take steps to preserve the functional integrity of the research commons.

A. The Challenge to Science: Formulating a Response to the Legal and Economic Pressures

The trends described above could elicit one of two types of responses. One is essentially reactive, in which the scientific community adjusts to the pressures as best it can without organizing a response to the increasing encroachment of a commercial ethos upon its upstream data resources. The other would require science policy to address the challenge by formulating a strategy that would enable the scientific community to take charge of its basic data supply and manage the resulting research commons in ways that preserved its public good functions without impeding socially beneficial commercial opportunities.

Under the first alternative, the research community can join the enclosure movement and profit from it.⁵²⁵ Thus, both universities and independent labora-

522. See discussion *supra* at Part III.B.2.d.(ii).

523. See discussion *supra* at Part III.B.1.

524. See discussion *supra* at Part III.B.2.a-c.

525. Cf. Boyle, *Second Enclosure Movement*, *supra* note 373.

tories or investigators that already transfer publicly funded technology to the private sector can also profit from the licensing of databases. In that case, data flows supporting public science will have to be constructed deal-by-deal with all the transaction costs this entails and with the further risk of bargaining to impasse.⁵²⁶ The ability of researchers to access and aggregate the information they need to produce discoveries and innovations may be compromised both by the shrinking dimensions of the public domain and by the demise of the sharing ethos in the nonprofit community, as these same universities and research centers increasingly see each other as competitors rather than partners in a common venture.⁵²⁷ Carried to an extreme, this competition of research entities against one another, conducted by their respective legal offices, could obstruct and disrupt the scientific data commons.

To avoid these outcomes, the other option is for the scientific community to take its own data management problems in hand. The idea is to reinforce and recreate, by voluntary means, a public space in which the traditional sharing ethos can be preserved and insulated from the commoditizing trends identified above.⁵²⁸ In approaching this option, the community's primary assets are the formal structures that support federally funded data and the ability of federal funding agencies to regulate the terms on which data are disseminated and used. The first programmatic response would look to the strengthening of existing institutional, cultural, and contractual mechanisms that already support the research commons, with a view to better addressing the new threats to the public domain identified above. The second logical response is collectively to react to new information laws and related economic and technical pressures by negotiating contractual agreements between stakeholders to preserve and enhance the research commons.⁵²⁹

As matters stand, the U.S. government generates a vast public domain for its own data through creative use of three instruments: intellectual property rights, contracts, and new technologies of communication and delivery. By long tradition, the federal government has used these instruments differently from the rest of the world. It waives its property rights in government-generated information, it contractually mandates that such information should be provided at the marginal cost of dissemination, and it has been a major proponent

526. Cf. Eisenberg, *Bargaining*, *supra* note 13, at 231-47.

527. See, e.g., Powell, *supra* note 13, at 266 (predicting turf wars between rival networks of partners).

528. Cf., e.g., Rai, *supra* note 33, at 144-51 (advocating norms-based approach to preserving use of biotech inventions for research purposes). What unites, or should unite, all these communities is a common understanding of the historical function of the public domain and a common need to preserve that function despite the drive for commoditization. See, e.g., Powell, *supra* note 13, at 265-66. Although legislators and entrepreneurs may take time to understand the threat that a shrinking public domain poses for the national system of innovation, the one group that is best positioned to appreciate that threat is the nonprofit research sector, whose dependence on the public domain remains a matter of everyday practice and vital concern. This sector is also the best positioned to take steps to respond to the threat by appropriate voluntary collective action.

529. Cf. UNIF. BIOLOGICAL MATERIALS TRANSFER AGREEMENT (UBMTA) (1995), available at <http://www.autm.net/UBMTA/intro.html> (last visited Jan. 18, 2003); Rai, *supra* note 33, at 113, nn.201-04 (discussing UBMTA).

and user of the Internet to make its information as widely available as possible. In other words, the U.S. government has deliberately made use of existing intellectual property rights, contracts, and technologies to construct a research commons for the flow of scientific data as a public good. The unique combination of these instruments is a key aspect of the success of our national research enterprise.⁵³⁰

Now that the research commons has come under attack, the challenge is not only to strengthen a demonstrably successful system at the governmental level, but also to extend and adapt this methodology to the changing university environment and to the new digitally networked research environment. In other words, universities, not-for-profit research institutes, and academic investigators, all of whom depend on the sharing of data, will have to stipulate their own treaties or contractual arrangements to ensure unimpeded access to, and unrestricted use of, commonly needed raw materials in a public or quasi-public space, even though many such institutions or actors may separately engage in transfers of information for economic gain.⁵³¹ This initiative, in turn, will require the federal government as the primary funder—acting through the science agencies—to join with the universities and scientific bodies in an effort to develop suitable contractual templates that could be used to regulate or influence the research commons.

Implementing our ideas would require nuanced solutions tailor-made to the needs of government, academia, and industry in general and to the specific exigencies of different scientific disciplines. The following sections describe our proposals for preserving and promoting the open availability of government-generated scientific data, and of government-funded and private-sector scientific data, respectively. We do not, however, develop detailed proposals for separate disciplines and sub-disciplines here, as these would require additional research and analysis.

B. Proposals for the Government Sector

To preserve and maintain the traditional public domain functions of government-generated data, the United States will have to adjust its existing policies and practices to take account of new information regimes and the growing pressures for privatization. At the same time, government agencies will have to find ways of coping with bilateral data exchanges with other countries whose governments choose to exercise intellectual property rights in their own data collections.

530. Mowery & Rosenberg, *supra* note 231, at 62 (stressing the interaction of federal and private R&D expenditures).

531. Our proposals go well beyond a “clearinghouse” approach to conflicting proprietary rights, which some have suggested. *See* discussion *supra* note 454. A clearinghouse does not deal with the positive externalities that could accrue from organizing worldwide flows of scientific data online, which is the real opportunity at stake.

1. Adjusting Domestic Policies and Practice

We do not mean to imply a need to totally reinvent or reorganize the existing universe in which scientific data are disseminated and exchanged. The opposite is true. As we have explained, a vast public domain for the diffusion of scientific data—especially government-generated data—exists and continues to operate, and much government-funded data emerging from the academic communities continues to be disseminated through well-established mechanisms.⁵³²

Facilities for the curation and distribution of government-generated data are well organized in a number of research areas. They are governed by long-established protocols that maintain the function of a public domain, and in most cases ensure open access (either free or at marginal cost) and unrestricted use of the relevant data collections. These collections are housed in bricks-and-mortar data repositories, many of which are operated directly by the government, such as the NASA National Space Science Data Center.⁵³³ Other repositories are funded by the government to carry out similar functions, such as the archives of the Hubble Space Telescope Science Institute at Johns Hopkins University.⁵³⁴

Under existing protocols, most government-operated or government-funded data repositories do not allow conditional deposits that look to commercial exploitation of the data in question. Anyone who uses the data deposited in these holdings can commercially exploit their own versions and applications of them without needing any authorization from the government. However, no such uses, including costly value-adding uses, can remove the original data from the public repositories. In this sense, the value-adding investor obtains no exclusive rights in the original data, but is allowed to protect the creativity and investment in the derived information products.

The ability of these government institutions to make their data holdings broadly available to all potential users, both scientific and other, has been greatly increased by direct online delivery. However, this potential is undermined by a perennial and growing shortage of government funds for such activities, by technical and administrative difficulties that impede long-term preservation of the exponentially increasing amounts of data to be deposited, and by pressures to commoditize data, which are reducing the scope of government activity and tend to discourage academic investigators from making unconditional deposits of even government-funded data to these repositories.⁵³⁵

The long-term health of the scientific enterprise depends on the continued operation of these public data repositories and on the reversal of the negative trends identified earlier in this article. Here, the object is to preserve and enhance the functions that government data repositories have always played,

532. See discussion *supra* at Parts II.A and II.B.

533. See NASA National Space Science Data Center, at <http://nssdc.gsfc.nasa.gov/> (last visited Jan. 10, 2003).

534. See Space Telescope Science Institute, at <http://www.stsci.edu/> (last visited Jan. 10, 2003).

535. See discussion *supra* at Parts III.B.1.

notwithstanding the mounting pressures to commoditize even government-generated data.

Implementing any recommendations concerning government-generated data will, of course, require adequate funding, and this remains a major problem. In most cases, however, it is not the big allocations needed to collect or create data that are lacking; it is the relatively small but crucial amounts to properly manage, disseminate, and archive data already collected that are chronically insufficient. These shortsighted practices deprive taxpayers of the long-term fruits of their investments in the scientific enterprise. Science policy must give higher priority to formulating workable measures to redress this imbalance than it has in the past.

Policy-makers should also react to the pressures to privatize government-generated research data by devising objective criteria for ascertaining when and how privatization truly benefits the public interest. At times, privatization will advance the public interest because the private sector can generate particular data sets more efficiently or because other considerations justify this approach. Very often, however, the opposite will be true, especially when the costs of generating the data are high in relation to known, short-term payoffs. Two recent National Research Council studies have attempted to formulate specific criteria for evaluating proposed privatization initiatives concerning scientific data.⁵³⁶ The science agencies should make the formulation of such criteria for different areas of research a top agenda item. In so doing, the agencies also need to analyze the results of past privatization initiatives with a view to assessing their relative costs and benefits.

Once the validity of any given privatization proposal has been determined by appropriate evaluative criteria, the next crucial step is to build appropriate,

536. The BITS OF POWER report set forth the following conditions for privatization of the scientific data distribution function:

- Can the distribution of data be separated easily from their generation?
- Is the scientific data set used by others beyond the research community?
- Is the potential market large enough to support several data distributors?
- Is it easy to discriminate prices or differentiate products between scientific users and other users? If this is possible, can low prices be mandated contractually for government-funded data for scientific users?
- Is it costly to separate the distribution of data to scientists from their distribution to other users, such as commercial users?

“If all of these questions can be answered ‘yes,’ then privatizing the distribution of scientific data should be an option to be considered.” BITS OF POWER, *supra* note 1, at 120-21.

Concerning the privatization of government data collection and product development functions in the environmental research context, a more recent National Research Council report recommended the following:

Before transferring government data collection and product development to private-sector organizations, the U.S. government should ensure that the following conditions will be satisfied: (1) avoidance of market conditions that will give any firms significant monopoly power; (2) preservation of full and open access to core data products; (3) assurance that a supply of high-quality information will continue to exist; and (4) minimized disruption to ongoing uses and applications.

RESOLVING CONFLICTS, *supra* note 8, at 87.

public-interest contractual templates into that deal, to ensure the continued operation of a research commons. The public research function is too important to be left as an afterthought. It must figure prominently in the planning stage of every legitimate privatization initiative precisely because the data would previously have been generated at public expense for a public purpose. After all, the process of privatization aims to shift the commercial risks and opportunities of data production or dissemination to private enterprise under specified conditions that promote efficiency and economic growth. However, that process should not pin the functions of the research enterprise to the success of any given commercial venture; it must not allow such ventures to otherwise compromise these functions by charging unreasonable prices or imposing contractual conditions unduly restricting public, scientific uses of the data in question.

There are two situations in which model contractual templates, developed through inter-agency consultations, could play a critical role. One is where data collection and dissemination activities previously conducted by a government entity are transferred to a private entity. The other is where the government licenses data collected by a private entity for public research purposes.⁵³⁷ In both cases, the underlying contractual templates should implement the following research-friendly legal guidelines:

- (1) A general obligation not to legally or technically hinder access to the data in question for nonprofit scientific research and educational purposes;
- (2) A further obligation not to hinder or restrict the reuse of data lawfully obtained in the furtherance of nonprofit scientific research activities;⁵³⁸ and
- (3) An obligation to make data available for nonprofit research and educational purposes on fair and reasonable terms and conditions, subject to impartial review and arbitration of the rates and terms actually applied, in order to avoid research disasters such as the Landsat deal in the 1980s.⁵³⁹

When the public data collection activity is transferred to the private sector, care must be taken to ensure that the private entity exercises any underlying intellectual property rights, especially some future database right, in a manner consistent with the public interest—including the interests of science. To this end, a model contractual template should also include a comprehensive misuse provision like that embodied in H.R. 1858.⁵⁴⁰

537. See, e.g., Commercial Space Act of 1998, Pub. L. No. 105-303, 112 Stat. 2843 (Oct. 28, 1998) (codified at 42 U.S.C. § 14701).

538. Moreover, even in those cases where government data are made available for private sector uses without any express transfer or contractual arrangements, agencies must give greater consideration to the need to preserve access to the original public data sets and avoid their de facto capture by a private entity.

539. See *supra* note 251 and accompanying text.

540. H.R. 1858 § 106(b), “Limitations on Liability,” provides as follows:

The larger principle is that, in managing its own public research data activities, the government can and should develop its own database law in a way that promotes science without unduly impeding commerce. This principle is not new; the government already has a workable information regime, as described in Part II. However, the government will need to adapt that regime to the pressures arising from the new high-protectionist legal environment to ensure that its agencies are consistently applying rational and harmonized public-interest principles. Otherwise, the traditional public domain functions of government-generated data could be severely compromised, an outcome that would violate the government's fiduciary responsibilities to taxpayers and raise conflicts of interest and questions concerning sham transactions.

2. Bridging the Gap with Foreign Law

The federal government will also have to continue to develop policies and procedures for dealing with data generated by foreign governments that commercialize their data and exploit all available intellectual property rights. Because international agreements concerning the exchange of scientific and technical information normally rely on national treatment clauses,⁵⁴¹ negotiated arrangements may be needed to bridge the differences between high and low-protectionist jurisdictions that could complicate international scientific cooperation.⁵⁴²

Ideally, arrangements with foreign governments should enable the United States to continue to waive intellectual property rights in government-generated data distributed abroad while requiring foreign governments similarly to waive intellectual property rights in government data disseminated in the United

(b) MISUSE—A person or entity shall not be liable for a violation of section 102 if the person or entity benefiting from the protection afforded a database under section 102 misuses the protection. In determining whether a person or entity has misused the protection afforded under this title, the following factors, among others, shall be considered:

- (1) the extent to which the ability of persons or entities to engage in the permitted acts under this title has been frustrated by contractual arrangements or technological measures;
- (2) the extent to which information contained in a database that is the sole source of the information contained therein is made available through licensing or sale on reasonable terms and conditions;
- (3) the extent to which the license or sale of information contained in a database protected under this title has been conditioned on the acquisition or license of any other product or service, or on the performance of any action, not directly related to the license or sale;
- (4) the extent to which access to information necessary for research, competition, or innovation purposes have been prevented;
- (5) the extent to which the manner of asserting rights granted under this title constitutes a barrier to entry into the relevant database market; and
- (6) the extent to which the judicially developed doctrines of misuse in other areas of the law may appropriately be extended to the case or controversy.

H.R. 1858; *see also* Dreier, *supra* note 68, at 295, 311-12; Ullrich, *supra* note 504.

541. *See, e.g.*, Agreement on Scientific and Technological Cooperation, U.S.-Vietnam, Nov. 17, 2000, cl. 2.2.

542. Reichman, *Database Protection in a Global Economy*, *supra* note 18, at 485-500.

States. Such a result would preserve the pure public domain approach to government-generated data that has long been official U.S. policy. However, European governments accustomed to commercializing their data reportedly have resisted this approach,⁵⁴³ presumably because they fear re-exports of the data back into their own more protected markets, or because they do not want to concede preferences to foreign users that they deny their own citizens. Requiring foreign governments to subscribe to the U.S. concept of an unconditional public domain for government-generated data may thus result in those governments disclosing considerably less data than they might under a two-tiered structure that conditionally allowed access to such data for nonprofit scientific and educational purposes while restricting its availability to the private sector.

While the U.S. tradition is squarely opposed to restricted uses of government-generated data, many European (and other) governments have subscribed to a different tradition.⁵⁴⁴ The E.C. Database Directive represents a powerful new thrust in that direction. It is worth reiterating that this Directive enables governments to exercise strong and potentially perpetual exclusive rights in *publicly* generated databases, without any mandated obligation to recognize public-interest exceptions.⁵⁴⁵

Some fifty countries either belonging to the European Union or having affiliated status are expected to adopt that model, and we believe that E.U. trade negotiators are seeking to impose it on other countries as part of regional trade agreements. If the United States fails to adopt a different, less protectionist database regime, founded on true unfair competition principles, the pressures for other countries to follow the E.U. model will become very great. Moreover, even if the United States adopts a significantly less protectionist database law, there will be pressures on the United States to protect data generated by foreign governments and made available to U.S. data centers despite the “no conditional deposit” rules that bind many of these centers. The United States, of course, will not be able to prevent foreign governments from commercially exploiting their public data in territories governed by the E.C. Database Directive. On the contrary, the fact that governments in the European Union themselves saw this Directive as a source of considerable income most likely disposed them favorably toward it, and this fatal attraction seems to be spreading.⁵⁴⁶

For these reasons, and despite the general undesirability of a two-tiered structure in the public sector, the United States must seek to persuade foreign governments that choose to exercise crown rights (both copyrights and *sui gene-*

543. Interview with Glenn Tallia, counsel for the National Oceanic and Atmospheric Administration, in Silver Spring, Md. (Aug. 20, 2000).

544. For an overview of the public information regimes of the E.U. member states, see Green Paper, *supra* note 46, at Annexe 1.

545. See discussion *supra* at Part III.B.2.d.i.

546. For example, the Republic of Korea is currently considering the adoption of a new *sui generis* database protection statute modeled on the E.C. Database Directive.

ris rights) under the E.C. Database Directive or its analogues to at least implement a conditional domain in their own countries, with a view to maximizing access for nonprofit research, educational, and other public-interest purposes.⁵⁴⁷ Obviously, the better result would be for the E.U. governments to renounce crown rights in public information altogether and to adopt the public domain policy of the U.S. government. Some efforts in this direction are in fact under way, but the outcome is highly uncertain.⁵⁴⁸ The overriding need to construct cooperative, worldwide open-data exchanges in support of public research and for addressing global problems—including environmental degradation, health, and the alleviation of poverty—provides a powerful mandate to achieve this result.

At the same time, there is a real danger that the European Union will continue to press intergovernmental organizations, as they have the World Meteorological Organization, to adopt two-tiered systems that deviate from established U.S. norms.⁵⁴⁹ The European Union may also be expected to press U.S. government agencies to conditionally protect the European Union's data in intergovernmental exchanges and thus, in effect, to institute a two-tiered approach for some purposes at U.S. data centers otherwise operating on a pure public domain basis. Similarly, the European Union may seek to persuade the U.S. government to retreat from its full and open data exchange policy in international scientific research programs.

These divergent pressures indicate that the rules applicable to intergovernmental exchanges of data may need to be revisited in the emerging high-protectionist legal environment. Senior representatives of the U.S. scientific community will have to make their voices heard in any such negotiations and argue the case for an international, open and cooperative public science regime to the greatest extent possible.⁵⁵⁰

C. Proposals for the Academic Sector

In putting forward our proposals concerning the preservation of a research commons for government-funded data, it is useful to follow the distinction between a zone of formally regulated data exchanges and a zone of informal data exchanges drawn earlier in this article. Consistent with our analysis in Part

547. See, e.g., Commission Proposal for a Council Directive on Public Access to Environmental Information, 2000 O.J. (C 337) 402; see also Commission Proposal for a Council Directive on the Reuse and Commercial Exploitation of Public Sector Documents, 2002 O.J. (C 227) 207.

548. *Id.*

549. For a discussion of the European Union's efforts for the World Meteorological Organization ("WMO") to adopt a two-tiered data distribution system, see Weiss & Backlund, *supra* note 7. For a statement of the two-tiered data policy that replaced the previous policy of full and open exchange at the WMO, see WORLD METEOROLOGICAL ORGANIZATION, EXCHANGING METEOROLOGICAL DATA: GUIDELINES ON RELATIONSHIPS IN COMMERCIAL METEOROLOGICAL ACTIVITIES (1996), available at <http://www.wmo.ch/web/pla/WMO837.pdf> (last visited Jan. 10, 2003).

550. This topic is under discussion already at the OECD. See Background Information on the Activities of the OECD Follow-up Group on Issues of Access to Publicly Funded Research Data, at <http://dataaccess.ucsd.edu/> (last visited Jan. 18, 2003).

II, we emphasize that the ability of government funding agencies to influence data exchange practices will be much greater in the formal than the informal zone.

1. Formally Regulated Data Exchanges

When no significant proprietary interests come into play, the optimal solution for government-generated data and data produced by government-funded research is a formally structured archival data center also supported by government. As discussed earlier, many such data centers have already been formed around large-facility research projects.⁵⁵¹ Building on the opportunities afforded by digital networks, it has now become possible to extend this time-tested model to highly distributed research operations conducted by groups of academics in different countries.

The traditional model entails a bricks-and-mortar centralized facility into which researchers deposit their data unconditionally. Besides academics, contributors may include government and even private sector scientists, but in all cases the true public domain status of any data deposited is usually maintained. Examples include the National Center for Biotechnology Information (“NCBI”),⁵⁵² which is directly operated by the National Institutes of Health, and the National Center for Atmospheric Research (“NCAR”),⁵⁵³ which is operated by a university consortium and funded primarily by the National Science Foundation (“NSF”).

A second, more recent model, enabled by improved Internet capabilities, also envisions a centralized administrative entity, but this entity governs a network of highly distributed smaller data repositories, sometimes referred to as “nodes.”⁵⁵⁴ Taken together, the nodes constitute a virtual archive whose relatively small central office oversees agreed technical, operational, and legal standards to which all member nodes adhere.⁵⁵⁵ Examples of such a decentralized network, which operate on a public domain basis, are the NASA Distributed Active Archive Centers under the Earth Observing System program⁵⁵⁶ and the NSF-funded Long Term Ecological Research Network.⁵⁵⁷

551. *See supra* Part II.A.

552. *See* discussion *supra* note 45.

553. *See* National Center for Atmospheric Research, at <http://www.ncar.ucar.edu/ncar/> (last visited Jan. 18, 2003).

554. Many networks of such distributed data nodes now exist. *See, e.g., supra* note 130; *see also* Planetary Data System, at <http://pds.jpl.nasa.gov/> (last visited Jan. 17, 2003).

555. *See* NATIONAL RESEARCH COUNCIL, PRESERVING SCIENTIFIC DATA ON OUR PHYSICAL UNIVERSE: A NEW STRATEGY FOR ARCHIVING THE NATION'S SCIENTIFIC INFORMATION RESOURCES 47-57 (1995) [hereinafter PRESERVING SCIENTIFIC DATA]; *see also supra* note 130.

556. *See* NASA's Earth Observing System, <http://eosps.gsfc.nasa.gov/> (last visited Jan. 17, 2003); NASA Distributed Active Archive Data Center Alliance, at <http://nasadaacs.eos.nasa.gov/> (last visited Feb. 3, 2003).

557. *See* LTER Net, *supra* note 130.

These virtual archives, known as “federated” data management systems,⁵⁵⁸ extend the benefits and practices of a centralized bricks-and-mortar repository to the outlying districts and suburbs of the scientific enterprise. They help to reconcile practice with theory in the sense that the investigators—most of whom are funded by government anyway—are encouraged to deposit their data in such networked facilities. The very existence of these formally constituted networks thus helps to ensure that the resulting data are effectively made available to the scientific community as a whole, which means that the social benefits of public funding are more perfectly captured and the sharing ethos is more fully implemented.

At the same time, some of the existing “networks of nodes” have already adopted the practice of providing conditional availability of their data: a feature of considerable importance for our proposals. By conditional availability we mean that the members of the network have agreed to make their data available for public science purposes on mutually acceptable terms, but they also permit suppliers to restrict uses of their data for other purposes, typically with a view to preserving their commercial opportunities.⁵⁵⁹

The networked systems thus provide prospective suppliers with a mix of options to accommodate deposits ranging from true public domain status to fully proprietary data that has been made available subject to rules the member nodes have adopted. The element of flexibility that conditional deposits afford makes these federated data management systems particularly responsive to the realities of present day university research in areas of scientific investigation where commercial opportunities abound.

a. Basic Recommendations

Our first proposition is that the government funding agencies should encourage unconditional deposits of research data, to the fullest extent possible, into both centralized repositories and decentralized network structures. The obvious principle here is that, because the data in question are government-funded, improved methods should be devised for capturing the social benefits of public funding, lest commercial temptations produce a kind of de facto free-riding at the taxpayers’ expense.⁵⁶⁰

When unconditional deposits occur in a true public domain environment removed from proprietary concerns, the legal mechanisms to implement these expanded data centers need not be complicated. Single researchers or small

558. PRESERVING SCIENTIFIC DATA, *supra* note 555, at 51-52 (describing the elements of a federated management structure).

559. An example of an international network that operates on the basis of conditional deposits is the Global Biodiversity Information Network (“GBIF”), headquartered in Denmark, which is substantially supported by U.S. government funding. *See* GBIF, *supra* note 130.

560. We use the term “free-riding” to suggest that privatization should not deprive the public of the full benefits that it paid for. At the same time, we recognize that entrepreneurs also pay taxes on their profits. *Cf.* Stephen Berry, Promoting Access to and Use of Not-for-Profit Sector Scientific and Technical Data—An Assessment of Legal and Policy Options, panel discussion at NRC Workshop, *supra* note 428, at Chapter 13.

research teams could contribute their data to centers serving their specific disciplines, with no strings attached other than measures to ensure attribution and professional recognition.⁵⁶¹ Alternatively, as newly integrated scientific communities organize themselves, they could seek government help in establishing new data centers or nodes that would accept unrestricted deposits on their behalf.⁵⁶² Private companies could also contribute to a true public domain model or organize their own variants of such a model; these practices should be encouraged as a matter of public policy.

If the unrestricted data were deposited in federal government sponsored repositories, existing federal information law and associated protocols would define the public access rights.⁵⁶³ The maintenance of public-interest data centers in academia, however, is problematic without government support. These data centers can become partly or fully self-supporting through some appropriate fee structures,⁵⁶⁴ but resort to a fee structure based on payments of more than the marginal cost of delivery quickly begins to defeat the public good and positive externality attributes of the system, even absent further use restrictions.

Leaving aside the funding issue, the deeper question that this first proposal raises is how the universities and other nonprofit research entities will resolve the potential conflict between the pressure to disclose and deposit their government-funded data and the valuable proprietary interests that are increasingly likely to surface in a high-protectionist intellectual property environment.⁵⁶⁵ One cannot ignore the risk that the viability and effectiveness of these centers could be undermined to the extent that the beneficiaries of government funding can resist pressure to further implement the sharing ethos and even to decline to deposit their research data because of their commercial interests.

Despite their educational missions and nonprofit status, universities and individual academics are both increasingly prone to regard their databases as targets of opportunity for commercialization.⁵⁶⁶ This tendency will become more pronounced as more of the financial burden inherent in the generation and management of scientific data is shouldered by the universities themselves

561. Academics are particularly concerned about receiving suitable attribution and recognition for their data-related activities. *See, e.g., Eisenberg, Proprietary Rights, supra* note 75, at 178. There is also evidence that one reason open-source software systems have succeeded is that they confer reputational (and other non-monetary) benefits on their participants. *See, e.g., Josh Lerner & Jean Tirole, Some Simple Economics of Open Source, 50 J. INDUS. ECON. 197 (2002).*

562. We note in this connection that many academics have self-organized mini-"data centers" through their websites with public domain functions, limited only by their technical and financial capabilities. Groups of academics can similarly construct more ambitious mini-centers, which become less elaborate versions of the government data model.

563. If, however, data centers are formed outside the scope of direct government control, the organizers and managers may need to reconstruct the public domain through general public use licenses to emulate the protocols that govern deposits of data in more traditional government-operated centers. *See* discussion *infra* Part IV.C.2.

564. *See, e.g., infra* notes 599-601 and accompanying text (discussing Swiss-PROT).

565. *Cf. Rai, supra* note 33, at 94-115.

566. *See, e.g., Powell, supra* note 13, at 265 ("[W]hat is striking is how actively universities and firms are seeking to privatize new information," not to profit from innovations, but because they "hope instead to profit from the supply of information or data analysis.").

or by cooperative research arrangements with the private sector. In this context, the universities are likely to envision split uses of their data and will prefer to make them available on restricted conditions. They will logically distinguish between uses of data for basic research purposes by other nonprofit institutions and purely commercial applications.⁵⁶⁷ Even this apparently clear-cut distinction might break down, moreover, if universities treat databases whose principal user base is other nonprofit research institutions as commercial research tools.

The point is that the universities may not want to deposit data in designated repositories, even with government support, unless the repositories can accommodate these interests, and the repositories could compromise their public research functions if they are held hostage to too many demands of this kind. The same potential situation exists for individual databases made available by universities (as opposed to their contributions to larger, multi-source repositories). This state of affairs will accordingly require still more creative initiatives to parry the economic and legal pressures on universities and academic researchers to withhold data.

With these factors in mind, our second major proposal is to establish a zone of conditionally available data in order to reconstruct and artificially preserve functional equivalents of a public domain. This strategy entails using property rights and contracts to reinforce the sharing norms of science in the nonprofit, trans-institutional dimension, without unduly disrupting the commercial interests of those entities that choose to operate in the private dimension.

To this end, the universities and nonprofit research institutions that depend on the sharing ethos, together with the government science funding agencies, should consider stipulating to suitable “treaties” and other contractual arrangements to ensure unimpeded access to commonly needed raw materials in a public or quasi-public space.⁵⁶⁸ From this perspective, one can envision the accumulation of shared scientific data as a community asset held in a contractually reconstructed research commons to which all researchers have access for purposes of public scientific pursuits.

One can further imagine that this public research commons exists in an ever-expanding “horizontal dimension,” as contrasted with the commercial operations of the same data suppliers in what we shall call the “vertical” or private dimension. The object of the exercise would be to persuade the government, as primary funder, to join with universities and scientific bodies in an effort to develop suitable contractual templates that could be used to regulate the research commons. These templates would ensure that data held in the quasi-public or horizontal dimension would remain accessible for scientific purposes and could not be removed or otherwise appropriated to the private or vertical

567. Cf. Eisenberg, *Bargaining*, *supra* note 13, at 242-43 (describing emergence of de facto two-tiered market in which scientists exchange research tools directly on minimal obligations, while technology transfer offices haggle over proprietary exchanges and delay research).

568. Cf. James Boyle, *A Politics of Intellectual Property: Environmentalism for the Net*, 47 *DUKE L.J.* 87 (1997); Rai, *supra* note 33, at 152 (advocating “concerted public and private action centered around existing norms to preserve the public domain”).

dimension. At the same time, these contractual arrangements would expressly contemplate the possibilities for commercial exploitation of the relevant data in the private or vertical dimension, and they would clarify the depositor's rights in that regard and ensure that the exercise of those rights did not impede or disrupt access to the horizontal space for research purposes.

b. Ancillary Considerations

In fashioning these proposals, we are aware that considerable thought has recently been given to the construction of voluntary social structures to support the production of large, complex information projects.⁵⁶⁹ Particularly relevant in this regard are the open source software movement that has collectively developed and managed the GNU/Linux Operating System⁵⁷⁰ and the Creative Commons organization,⁵⁷¹ which seeks to encourage authors and artists to conditionally dedicate some or all of their exclusive rights to the public domain. In both these pioneering movements, agreed contractual templates have been experimentally developed to reverse or constrain the exclusionary effects of strong intellectual property rights.

The open source model adopted by the software research and related communities relies on existing legal regulatory regimes to create a social space devoted to producing freely available and modifiable code.⁵⁷² Under the GNU/Linux operating system, components of the cooperatively elaborated structure are protected by intellectual property rights, in this case copyrights, and by licensing agreements, but these legal mechanisms are used to enforce the sharing norms of the open source community.⁵⁷³ Standard-form licensing agreements are formulated "to use contractual terms and property rights to create social conditions in which software is produced on a model of openness

569. See, e.g., Benkler, *Coase's Penguin*, *supra* note 198; David McGowan, *Legal Implications of Open-Source Software*, 2001 U. ILL. L. REV. 241, 245 (2001); see also LAWRENCE LESSIG, *THE FUTURE OF IDEAS: THE FATE OF THE COMMONS IN A CONNECTED WORLD* (2001); Dan Burk, *Open Source Genomics*, 8 B.U. J. SCI. & TECH. L. 254 (2002).

570. See, e.g., McGowan, *supra* note 569. The Open Source Movement is also known as the Free Software Movement. On this terminology, see SAM WILLIAMS, *FREE AS IN FREEDOM: RICHARD STALLMAN'S CRUSADE FOR FREE SOFTWARE*, Ch. 11 (2002), available at <http://www.oreilly.co/openbook/freedom/ch11.html> (last visited Dec. 15, 2002); see also Free Software Foundation, *Why 'Free Software' is Better than 'Open Source'*, at <http://www.gnu.org/philosophy/free-software-for-freedom.html> (last visited Jan. 10, 2002).

571. See Creative Commons, at <http://www.creativecommons.org/> (last visited Jan. 18, 2003).

572. McGowan, *supra* note 569, at 244.

573. Under the Free Software approach, the archtypical "copyleft license" is the GPL or GNU Public License, by which the copyright owner grants the user the right to copy, modify and distribute the licensed software without having to get permission or pay any license fee to the owner. "[I]f you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have." GNU GPL, available at <http://www.gnu.org/licenses/gpl.html> (last modified July 15, 2001). According to Janet Hope, a Ph D. candidate at the Australian National University, open source was intended as a market strategy to make the "free software" concepts attractive to people in the business community. Janet Hope, *Open Source Biotechnology*, paper presented to the workshop on Science, Intellectual Property and Open Domains, REGNET, Intellectual Property Institute of Australia, Canberra, Australia (Dec. 2, 2002).

rather than exclusion.”⁵⁷⁴ Under these licenses, “code may be freely copied, modified, and distributed,” but only if the modifications (derivative works) are distributed under these terms as well.⁵⁷⁵ Property rights are “held in reserve to discipline possible violations of community norms.”⁵⁷⁶ The end result, as Professor McGowan recently observed, is not a true commons, but resembles a commons because of the “low cost of copying and using code combined with . . . broad grants of the relevant licenses.”⁵⁷⁷

For present purposes, the most relevant lesson to be drawn from the open source model is the possibility for participants in networks of nodes or other data sharing arrangements to dedicate holdings protected by IPRs to the relevant scientific community itself, which would hold the collective asset in a kind of trust to which all members of the community have access. In effect, the members of such a community would use any exclusive rights granted by intellectual property laws to exclude exclusivity itself. While the collective asset—in this case, typically, a database—and its components could be routinely made available for commercial applications,⁵⁷⁸ subject to additional terms and conditions that would have to be negotiated, the general public licenses supporting the collective asset would prevent any users from appropriating either the entire asset or its components from the quasi-public or horizontal space in which it was collectively managed for public research purposes.

The second model of particular interest, the Creative Commons, facilitates public access to copyrighted literary and artistic works by devising a set of standard-form contractual templates any author can digitally adopt.⁵⁷⁹ Once adopted, these contractual grants permit anyone to make certain uses of the

574. McGowan, *supra* note 569, at 243. The movement makes some effort to downplay their reliance on technical rules of contract law as such. See, e.g., Eben Moglen, *Enforcing the GPL*, LINUXUSER & DEVELOPER, Sept./Oct. 2001, at 66.

575. McGowan, *supra* note 569, at 242.

576. *Id.* at 244.

577. *Id.*

578. Open source software makes money for developers in two principal ways. First, all innovations made after the initial release are automatically available to the original developers at no cost, which can be equivalent to having an enormous, unpaid R&D department. Hope, *supra* note 573. The second type of income stream comes from providing service and support, “which is always a large part of the market in any high tech industry.” *Id.*

579. The Creative Commons offers four standard templates:

Attribution. You let others copy, distribute, display, and perform your copyrighted work—and derivative works based upon it—but only if they give you credit. . . .

Noncommercial. You let others copy, distribute, display and perform your work—and derivative works based upon it—but for noncommercial purposes only. . . .

No Derivative Works. You let others copy, distribute, display and perform only verbatim copies of your work, not derivative works based upon it. . . .

Share Alike. You allow others to distribute derivative works only under a license identical to the license that governs your work.

Licenses Explained, available at <http://www.creativecommons.org/learn/licenses> (last visited Jan. 18, 2003).

The service also offers to help people dedicate their work to the pure public domain, with “no rights reserved,” although it is not provided as one of the four standard options. See Creative Commons, *supra* note 571.

protected works, which are then digitally encoded, so that the search engines of would-be users can register them and thus facilitate the uses in question.⁵⁸⁰ This technique seems particularly relevant to the goal of linking highly distributed data holders in virtual archives by digital means, as is further discussed below.⁵⁸¹

Although neither of these models were developed with the needs of public science in mind, both provide helpful examples of how universities, federal funding agencies, and scientific bodies might contractually reconstruct a research commons for scientific data that could withstand the legal, economic, and technological pressures on the public domain identified in this article. In what follows, we draw on these and other sources to propose the contractual regulation of government-funded data in two specific situations: (1) when government-funded, university-generated data are licensed to the private sector, and (2) when such data are made available to other universities for research purposes.

c. Licensing Government-Funded Data to the Private Sector

In approaching this topic, one must consider that the production of scientific databases in academia is not always dominated by activities funded by the federal government. It may also entail funding by universities themselves, foundations, and the private sector. While funding from these non-government sources seems likely to grow in the future, especially if Congress adopts a database protection right, the government's role in funding academic data production will nonetheless remain a major factor, at least in the near term (though its role will vary from project to project). As discussed in Part II, this presence gives the federal funding agencies unique opportunities to influence the data-sharing policies of its beneficiary institutions.⁵⁸²

Ideally, funders and universities would agree on the need to maintain the functions of a public domain to the fullest extent possible, to provide open access to data for nonprofit research activities, and to encourage efficient technological applications of available data. At the same time, technological applications and other opportunities for commercial exploitation of certain types of databases will push the universities to enter into private contractual transactions that, if left totally unregulated, could adversely affect the availability of the relevant data for public research purposes.⁵⁸³ The reconciliation of the conflict between enhancing the public research interests and freedom of contract will require carefully formulated policies and institutional adjustments.⁵⁸⁴

580. *Id.*

581. See discussion *infra* at Part IV.C.2.

582. See discussion *supra* Part II.B.1.a.

583. See, e.g., Eisenberg, *Bargaining*, *supra* note 13, at 229 (“Institutions tend to be high-minded about the importance of unfettered access to the research tools that they want to acquire from others, but no institution is willing to share freely the materials and discoveries from which they derive significant competitive advantage.”).

584. Cf. Dan L. Burk, *Lex Genetica: The Law and Ethics of Programming Biological Code*, 4 ETHICS & INFO. TECH. 109, 109-11, 113-18 (2002).

Assuming the existence of sufficient funds, the maximum availability of academic data for research purposes is assured if those data have been deposited in the public data centers.⁵⁸⁵ To the extent that agencies successfully encourage academics and their universities to deposit government-funded data into either old or new repositories established for this purpose, the research-friendly policies of these centers should automatically apply. As long as these policies are not themselves watered down by commercial and proprietary considerations, they should generally immunize the research function from conflicts deriving from private transactions.⁵⁸⁶

However, the universities or their academics may very well balk at contributing commercially valuable data to these repositories unless they retain some degree of autonomy to negotiate the terms of their private transactions and impose restrictions on the uses of the data deposited for commercial purposes. This raises two important questions. The first concerns the willingness of data centers themselves—whether of the bricks-and-mortar variety or networks of nodes—to accept conditional deposits that impose restrictions on use for certain purposes in the first place. The second question, closely tied to the first, concerns the extent to which federal funding agencies should further seek to define and influence the relations between universities and the private sector to protect the public research function—especially when the data in question have not been deposited in an appropriate repository or when they have been so deposited but the repository permits conditional deposits.

i. Key Questions

Regarding the first of these questions, we previously observed that the emerging network of nodes model is more likely to accommodate conditional deposits or availability than are the traditional centralized data centers.⁵⁸⁷ Nevertheless, the practice remains controversial in scientific circles in that it deviates from the traditional norm of full and open access. For present purposes, we simply state our view that the possibilities for maximizing access to scientific data for public nonprofit research will not be fully realized in a highly protectionist legal and economic environment unless the scientific community agrees to experiment with suitably regulated conditional deposits.⁵⁸⁸

The second question, concerning the need to regulate the interface between universities and the private sector with regard to government-funded data, acquires important contextual nuances when viewed in the light of the policies and practices that currently surround the Bayh-Dole Act and related legislation. The Bayh-Dole Act encourages universities to transfer the fruits of federally

585. See, for example, the data centers mentioned in *supra* note 45.

586. Moreover, there are likely to be significant positive externalities when private companies can freely use the public research data. See, e.g., Powell, *supra* note 13, at 263-65.

587. See *supra* Part IV.C.1.

588. Cf. Patrinos & Drell, *supra* note 23, at 11.

funded research to the private sector by means of the patent system.⁵⁸⁹ In a somewhat similar vein, federal research grants and contracts allow researchers to retain copyrights in their published research results. By extension, the same philosophy could apply to databases produced with federal funding, especially if Congress were to adopt a *sui generis* database protection right, with incalculably negative results unless steps were taken to reconcile the goals of Bayh-Dole with the dual nature of data as both an input and an output of scientific research and of the larger system of technological innovation.⁵⁹⁰

It would also be a mistake for the science policy establishment to wait for the enactment of database legislation before considering the implications of blindly applying the spirit of Bayh-Dole to any database law that Congress might adopt. Because databases differ significantly from either patented inventions or copyrighted research results, policy-makers should anticipate the advent of some database legislation and address the problems it may cause for science—particularly with regard to government-funded data. Special consideration must be given to how the power to control uses of scientific data after publication would be exercised once a database protection law was enacted.

We do not mean to question the underlying philosophy or premises of Bayh-Dole,⁵⁹¹ which has produced socially beneficial results. Its very success, however, has generated unintended consequences and raised new questions that require careful consideration.⁵⁹² In advocating a program for a contractually reconstructed research commons, one of our explicit goals is, indeed, to ensure that academics and their universities benefit from new opportunities to exploit research data in an industrial context. This goal reflects the policies behind Bayh-Dole.⁵⁹³ At the same time, it would hardly be consistent with the spirit of Bayh-Dole to allow the commercial partners of academic institutions to dictate the terms on which government-funded data are made available for purposes of nonprofit scientific research.

On the contrary, a real opportunity exists for government funding agencies and universities to develop agreed contractual templates that would apply to commercial users of government-funded data in general. In effect, the public scientific community would thus develop a database protection scheme of its own that would override the less research-friendly provisions of any *sui generis*

589. Even without a database right, Bayh-Dole prods universities to exploit data as part of the transfer of technology to the private sector. *See also* Rai, *supra* note 33, at 97, 109-15.

590. Of course, Bayh-Dole does not, and need not, apply to database protection rights as such. *See* 35 U.S.C. § 200 (2000) (limiting policy to use of the patent system). But universities would still remain unrestricted owners or assignees of government-funded databases, who could make their own rules, while Bayh-Dole already accustoms them to the commercial exploitation of their intellectual property rights. *See, e.g.,* Eisenberg, *Bargaining*, *supra* note 13; Powell, *supra* note 13, at 255 (stressing the new role of universities as both creators and retailers of intellectual property).

591. 35 U.S.C. § 200 (2000) (“Policy and Objectives”).

592. *See* Eisenberg, *Public Research*, *supra* note 120; Bar-Shalom & Cook-Deegan, *supra* note 121; *see also* Rai & Eisenberg, *supra* note 117.

593. 35 U.S.C. § 200 (listing, *inter alia*, the goal “to promote colatoration between commercial concerns and nonprofit organizations, including universities”).

regime that Congress might adopt. In so doing, the scientific community could also significantly influence the data-licensing policies and practices of the private sector, before that sector ends up influencing the data-licensing practices of university technology transfer offices.⁵⁹⁴

ii. Value-Adding Uses and Management Costs

If one takes this proposal seriously, a capital point of departure would be to address the problem of follow-on applications, which has greatly perturbed the debate about database protection in general. The critical role of data as input into the information economy weighs heavily against endowing database proprietors with an exclusive right to control follow-on applications. This principle becomes doubly persuasive when the government itself has defrayed the costs of generating the data in question, in which case an exclusive right to control value-added applications takes on a cast of reverse free-riding.⁵⁹⁵

One solution is to allow second-comers to extract and use data from any given collection freely for bona fide value-adding purposes in exchange for adequate compensation of the initial investor based on an expressly limited range of royalty options.⁵⁹⁶ If the rules developed by universities and funding agencies imposed this kind of “compensatory liability” regime on follow-on applications of government-funded academic data, in lieu of any statutorily created exclusive right, there is reason to believe it could significantly advance both technological development and the larger public interest in access to scientific data.⁵⁹⁷

Universities and funding agencies could also adopt clauses similar to those proposed above in the context of government-generated data,⁵⁹⁸ including a general prohibition against legally or technically hindering access to any database built around government-funded data for purposes of nonprofit scientific research. Clauses that prohibit private partners from hindering the reuse of data in the construction of new databases to address new scientific research objectives seem particularly important, as are clauses requiring private partners to license their commercial products on fair and reasonable terms and conditions. Also desirable are clauses forbidding misuse of any underlying IPRs and establishing guidelines that courts should apply in evaluating specific claims of misuse.

Moreover, when considering relations with the private sector, attention should be given to the high cost of managing and archiving data holdings for scientific purposes and the possibilities of defraying some of this cost through

594. Cf. Rai, *supra* note 33, at 114 (reporting instances in which the “[r]esidual norms of academic research may even have had some influence on the conduct of [biotech] industry actors”).

595. In the sense that it makes the public pay twice for some of the social benefits that public funding was designed to cover.

596. Cf. Reichman, *Of Green Tulips*, *supra* note 164; Reichman & Uhler, *Database Protection*, *supra* note 4; Reichman, *Database Protection in a Global Economy*, *supra* note 18, 462-63, 479-80.

597. Cf. Powell, *supra* note 13, at 264 (fearing exclusive license mentality that, if applied to Cohen-Boyer patent on recombinant DNA technology, would have retarded progress in biotechnology).

598. See *supra* Part II.B.1.

commercial exploitation. While government support ought to increase, especially as the potential gains from a horizontal e-commons become better understood, the cost of data management will also increase with the success of the system. For this reason, universities may want to levy charges against users in the private sector or the vertical dimension, in order to help defray the cost of administering operations in the horizontal domain and to make this overall approach more economically feasible.

One controversial example of an attempt to supplement the data management costs of a government-funded database is provided by the Swiss-PROT Protein Knowledgebase, a university-administered entity that collects and curates data concerning protein sequences contributed by academic and corporate researchers from various countries.⁵⁹⁹ The highly specialized data in this collection are reviewed, annotated, and made available online under a conditional arrangement that operates largely on an honor system.⁶⁰⁰ While nonprofit uses are allowed gratis, uses by private firms are licensed on an annual subscription fee basis that differentiates by the size of the corporate user (and, presumably, its ability to pay). Payments are made directly to the Swiss-PROT management entity and, reportedly, there has been minimum evasion of the rules thus far.⁶⁰¹

Another interesting example is the ultimately unsuccessful attempt to negotiate a public-private partnership to manage a Human Mutations Database, which would have integrated university-generated data from around the world concerning genomic mutations into a single, openly available database.⁶⁰² The motivating idea was that a private firm, Incyte, would put up the funds to organize the database and make the data openly available to all users online, on the condition that all of the relevant data would have been channeled through Incyte alone and not its competitors. Incyte expected to gain lead time advantages from early access to the data and also to benefit from traffic-building effects on its website, but there was to be no fee charged for use of the mutations data as such. In the end, however, it proved impossible to coordinate the disparate interests of the participating entities, and the project was aborted before its feasibility could be tested.⁶⁰³

In evaluating these experiments, one may view the so-called open access system adopted by Swiss-PROT with a degree of skepticism. First of all, the relevant user community is so small and tightly knit, and so accurately monitored by tracking the electronic footprints of those who access the database, that non-compliance would pose unacceptable costs in loss of reputation, peer pressure, and possible denial of privileges. Second, the administrators are able

599. See Swiss-PROT, at <http://www.ebi.ac.uk/swissprot/> (last visited Dec. 4, 2002).

600. See <http://www.ebi.ac.uk/swissprot/information/information.html> (last visited Dec. 4, 2002).

601. Reported through an informal discussion at a European Science Foundation, Funding Agencies Workshop on Public Domain of Digital Research Data, Strasbourg, France, Oct. 15, 2002.

602. Maurer, *Inside the Anticommons*, *supra* note 465.

603. *Id.*

tacitly to rely on the default rules, valid against the world, that derive from the E.C. Database Directive, which effectively holds every user who fails to comply with the posted conditions of access, extraction, and reuse liable for infringement.

Swiss-PROT nonetheless exemplifies one potential use of an agreed contractual template, and it anticipates techniques the Creative Commons initiative has recently further refined. It illustrates that, even in the presence of a high-protectionist intellectual property regime, contractual templates can be fashioned to promote the research commons without unduly obstructing commercial opportunities (although there are questions about Swiss-PROT's specific practices in this regard).⁶⁰⁴

The Swiss-PROT example also supports an otherwise intuitive inference that, under certain conditions, an intellectual property right in collections of data can encourage—rather than discourage—disclosure for scientific purposes by reducing the risk of free-riding appropriations. This comes as no surprise to the authors of this article, given that we have elsewhere advocated the adoption of a minimalist database protection regime, sounding in unfair competition law (liability rules) rather than in exclusive property rights.⁶⁰⁵ A database protection statute based on true unfair competition principles could close any demonstrable gaps in existing law with acceptable social costs and would provide a moderate, alternative model for other countries to consider.⁶⁰⁶ It should be clear, however, that the stronger the underlying intellectual property right, the more necessary it becomes to devise suitable contractual templates regulating relations between universities and the private sector (and inter-university relations themselves), with a view to ensuring the smooth operation of a contractually reconstructed research commons.

Moreover, complexities and coordination problems are likely to arise when the data in question are of interest to a much broader and more heterogeneous non-expert user base than in the two examples above. In this situation, more refined contractual templates could reduce both friction and transaction costs due to strategic behavior, by, for example, differentiating categories of users who may be denied access for specified lead time intervals; regulating the timing of competing or derivative publications; prospecting the possibilities of strategic cross-licenses for certain purposes; and even ensuring that grant-back, reach-through, and other clauses, sometimes appropriate in the private sector,

604. It should be noted that Swiss-PROT's default provision for nonprofit users was not to modify the database for any purpose, unless expressly allowed for "a valid scientific or technical reason" depending on "how the information will be presented and distributed." Maurer, *Promoting and Disseminating Knowledge*, *supra* note 259, at 47. This was a scientifically serious limitation. The NCBI, which previously incorporated Swiss-PROT data into its own "Reference Sequence" and "Predicted Genes" databases, stopped doing so when Swiss-PROT added this encumbrance and eventually replaced the Swiss-PROT data with data from other sources. *Id.*

605. See generally Reichman & Uhler, *Database Protection*, *supra* note 4; Reichman & Samuelson, *supra* note 18.

606. See generally Reichman, *Database Protection in a Global Economy*, *supra* note 18, at 493-96.

are not allowed to disrupt public research.⁶⁰⁷ However, the more complicated the situation becomes and the greater the degree of coordination required, the more likely worthwhile pooling initiatives will never get off the ground—as occurred with the human mutations database project itself.⁶⁰⁸

In addition, if such a project were successfully launched and the data in question became potentially of value to a broad user base, a further enforcement problem might arise due to the potential for leakage of data, supplied at preferential prices, to research users in ways that could damage the interests of users in the vertical dimension. It will be recalled that, on the horizontal plane, the option to charge for research uses (when otherwise unavoidable) is intended to entail a corresponding burden to positively discriminate in favor of science and its research goals.⁶⁰⁹ This need for price discrimination favoring research uses correspondingly requires that the difficult problem of leakage be addressed. Any solution here would certainly benefit from congressional enactment of a minimalist database protection right.⁶¹⁰

When administrative complexities appear particularly daunting, the better solution may be for participating entities to deposit their data with a designated, external administrative agency or service charged with the tasks of negotiating, formulating, and implementing the general public licenses or agreed contractual templates. These operations should remain subject to the guidance, governance, and oversight of the participating universities, government funding agencies, and other affected institutions. We also envision the need for mediation, arbitration, and dispute settlement facilities, which could be appropriately located within any oversight group that might be established.⁶¹¹

Finally, care must be taken to reduce friction between the scientific data commons as we envision it and universities' patenting practices under the Bayh-Dole Act. For example, any agreed contractual templates might have to allow for deferred release of data, even into repositories operating as a true public domain, at least for the duration of the one-year novelty grace period during which relevant patent applications based on the data could be filed.⁶¹² Other

607. Cf. Eisenberg, *Bargaining*, *supra* note 13, at 226-31 (evidencing disruptive effects of such clauses at present in regard to biotech research tools).

608. See, e.g., Rai, *supra* note 34, at 129-35; Maurer, *Inside the Anticommons*, *supra* note 465.

609. This practice is, of course, further restrained to the extent that the U.S. government provides the bulk of the data in the true public domain, which intrinsically restricts the amount of data available for providers who seek to opt into the conditional domain, with price-discriminated operations, along the horizontal research plane in addition to commercial operations at full rates along the vertical axis. See *supra* Part II.A.1.

610. The leakage problem might require administrators of a scientific e-commons to adopt and apply "access management systems" even though strong "digital rights management techniques" would be inappropriate in the interests of implementing and enforcing the community's norms. Institutional users are unlikely to disregard contractual access terms. See Peter Eckersley et al., *Neuroscience Data and Tool Sharing: A Legal and Policy Framework for Neuroinformatics*, 1 NEUROINFORMATICS 8-10 (forthcoming 2003).

611. Cf. *id.* at 10-11 (discussing a collection society model for similar purposes). A logical organizational locus for such operations would be the professional scientific societies working within the framework of the American Association for the Advancement of Science.

612. See 35 U.S.C. § 102 (2000); see also 35 U.S.C. § 111(b) (provisional patent applications).

measures to synchronize the operations of the e-commons with the ability of universities to commercialize their holdings under Bayh-Dole would have to be identified and carefully addressed. We also note that there is an interface between our proposals for an e-commons for science and antitrust law,⁶¹³ which would at least require consultation with the FTC and might also require enabling legislation. A detailed analysis of these issues lies beyond the scope of this article.

In sum, to successfully regulate relations between universities and the private sector in the United States, where most of the scientific data in question are government-funded (if not government-generated), considerable thought must be given to devising suitable contractual templates that universities could use when licensing such data to the private sector. These templates, which should aim to promote the smooth operation of a research commons and facilitate general research and development uses of data as inputs into technological development, could themselves constitute a model database regime that optimally balances public and private interests in ways that any federally enacted law might not. To succeed, however, these templates must be acceptable to the universities, the funding agencies, the broader scientific community, and the specific disciplinary sub-communities—all of whom must eventually weigh in to ensure that academics themselves observe the norms that they would thus have collectively implemented.

In so doing, the participating institutions could avoid a race to the bottom in which single universities might otherwise trade away more restrictions on open access and research to attract more and better deals from the private sector. Unless science itself takes steps of this kind, there is a serious risk that, under the impetus of Bayh-Dole, the private sector will gradually impose its own database rules on all government-funded data products developed with their university partners.⁶¹⁴

d. Inter-University Licensing of Scientific Data

Whatever the merits of our proposals for regulating transfers of scientific data from universities to the private sector, the need for science policy to regulate inter-university transfers of such data seems irrefutable. In this context, most of the data is generated for public scientific purposes and at public expense, and the progress of science depends on continued access to, and further applications of, such data. Not to construct a research commons that could withstand the pressures to privatize government-funded data at the inter-university level would thus amount to an indefensible abdication of the public trust by encumbering nonprofit research with high transaction and exclusion costs.⁶¹⁵

613. Cf. Dreier, *supra* note 68, at 311-12; Ullrich, *supra* note 504, at 367-98.

614. Cf. Rai, *supra* note 33, at 111-12 (reporting instances in which this has occurred regarding biotech patents).

615. See, e.g., Eisenberg, *Bargaining*, *supra* note 13, at 234 (“When progress in research depends on the relatively unfettered flow of low value exchanges of information and materials among scientists, a

All the same, implementing this task poses very difficult problems that are likely to exacerbate the conflicts of interests between the open and cooperative norms of science and the quest for additional funding sources we previously identified.⁶¹⁶

i. Policy Considerations

One may note at the outset that these conflicts of interest are rooted in the Bayh-Dole approach to the transfer of technology itself. This legislative framework stimulates universities to protect federally funded research results through intellectual property rights and to license those rights to the private sector for commercial applications. If Congress enacted a strong database-protection law, it could extend Bayh-Dole to this new intellectual property right. In such a case, Bayh-Dole would simply pass the relevant exclusive rights to extract and reutilize collected data straight through the existing system to the same universities and academic researchers who now patent their research results and would thus end up owning all the government-funded data they had generated. Even without such legislation, nothing impedes the universities from commercially exploiting protected databases in the spirit of Bayh-Dole, subject to any exceptions or immunities favoring research that a database law may have codified.

Moreover, the Bayh-Dole legislation makes no corresponding provision for beneficiary universities to give differential and more favorable treatment to other universities when licensing patented research products. On the contrary, there is evidence that in transactions concerning patented biotech research tools, at least, universities have viewed each other's scientists as a target market, in the exploitation of which they have virtually the same commercial interests as private producers of similar tools for scientific research.⁶¹⁷ Inter-university deals have accordingly been constructed on a case-by-case basis, often with considerable difficulty, by technology transfer offices normally striving to maximize all their commercial opportunities.⁶¹⁸

Without any agreed restraints on how universities are to deal with collections of data in which they had acquired statutorily conferred ownership and exclusive exploitation rights, their technology transfer offices could simply treat databases like patented inventions—despite the immensely greater impact this could have on both basic and applied research. In this milieu, reliance on good faith accommodations hammered out by the respective technology transfer offices⁶¹⁹ would, at best, make inter-university exchanges resemble the complicated transactions that already characterize relations between highly distributed

proliferation of intellectual property claims . . . may impose transaction costs that consume the gains from exchange.”).

616. See discussion *supra* Part III.B.1.b.

617. See Eisenberg, *Public Research*, *supra* note 120; Rai & Eisenberg, *supra* note 117 at 297-99. *But see id.* at 300-06 (countervailing efforts to preserve benefits of research commons).

618. See, e.g., Eisenberg, *Bargaining*, *supra* note 13, at 228-48.

619. See COHEN & MERRILL, *supra* note 452.

laboratories and research teams in the zone of informal exchanges of scientific data.⁶²⁰ All the vices of that zone would soon be imported into the more formal zone of inter-university relationships. At worst, this would precipitate a race to the bottom as universities tried to maximize their returns from these rights, in which case some technology transfer offices could be expected to contractually override any modest research exceptions a future database law might have codified.⁶²¹

At the same time, the Bayh-Dole legislative framework may itself suggest an antidote for resolving these potential conflicts of interest, or, at least, a sound point of departure for addressing them. The Act explicitly recognizes that the public interest in certain patented inventions may outweigh the benefits usually anticipated from private exploitation under exclusive property rights. In such cases, it authorizes the government to impose a compulsory license or otherwise exercise “march-in” rights and take control of the invention it has paid to produce.⁶²² In fact, these public-interest adjustments have never successfully been exercised in practice,⁶²³ and on the one known occasion when they were invoked, the government encountered stiff and questionable resistance from a major university.⁶²⁴

Nevertheless, the principle (if not the actual practice) behind these provisions presents a platform on which universities and federal funding agencies can build their own mutually acceptable arrangements to promote their common interest in full and open access to government-funded collections of data. Our goal, indeed, is to persuade them to address this challenge now, before a database protection law is enacted, by examining how to ensure the smooth and relatively frictionless exchange of scientific data between academic institutions, regardless of any exclusive property right they may eventually acquire and notwithstanding any other commercial undertakings with the private sector they may pursue. Absent such a proactive approach, we fear a slow unraveling of the traditional sharing norms in the inter-university context and an inevitable race to the bottom.

ii. Structuring Inter-University Data Exchanges

Because the issues under consideration here pertain to uses of government-funded data produced by academics for university-sponsored programs, one looks to full and open access as the optimal guiding principle and to the sharing

620. See discussion *supra* Part II.B.2.

621. At present, there are no such proposed exceptions of any real value in the pending U.S. exclusive rights model. See *supra* Part III.B.2. Even if this were to change, that model allows database producers to override such exceptions by contract, and the usual status of sole-source provider on many scientific niche markets provides the market power to do so.

622. 35 U.S.C. § 203 (2000); see, e.g., Frischmann, *supra* note 70, at 402-03.

623. In the private sector, the use of “march-in rights” could raise the specter of uncertainty and hamper investment. But this risk is of secondary importance in the inter-university environment.

624. In the case of CellPro, Johns Hopkins University successfully opposed a competitor licensee’s request for a compulsory license that would have brought an effective cancer diagnosis tool to the market years ahead of its own patented invention. See Bar-Shalom & Cook-Deegan, *supra* note 121.

norms of science as the foundation of any arrangement governing inter-university licensing of data. On this approach, the government-funded data collections held by universities would be viewed as a single common resource for inter-university research purposes. The operational goal would be to nurture and extend this common resource within a horizontally linked administrative framework that facilitated every university's public research functions, without unduly disrupting commercial relations with the private sector that some departments of some universities will undertake in the vertical dimension.

To achieve this goal, universities, funding agencies, and interested scientific bodies would have to negotiate an acceptable legal and administrative framework, analogous to a multilateral pact, that would govern the common resource and provide day-to-day logistical support. Ideally, the participating universities or their designated agents would operate as trustees for the horizontally constructed common resource, much as the Free Software Foundation does with the GNU system.⁶²⁵ In this capacity, the trustees would assume responsibility for ensuring access to the holdings on the agreed terms and for restraining deviant uses that violate those terms or otherwise undermine the integrity of the commons. The full weight of the federal granting structure could then be made to support these efforts by mandating compliance with agreed terms and directly or indirectly imposing sanctions for non-compliance.

Alternatively, a less formal administrative structure could be built around a set of agreed contractual templates regulating access to government-funded data collections for public research purposes. On this approach, the participating universities would retain greater autonomy, there would be less need for a fully fleshed out multilateral pact, and the monitoring and other transaction costs might be reduced. The Swiss-PROT arrangement discussed earlier⁶²⁶ provides some elements of an approach that could be adapted in a highly protectionist environment to promote certain inter-university data exchanges along these lines.

In a less than perfect world, however, there are formidable obstacles standing in the way of a negotiated commons project, over and above inertia, that would have to be removed. Initially, the very concept of an e-commons needs to be sold to skeptical elements of the scientific community whose services are indispensable to its development. Academic institutions, science funders, the research community, and other interested parties must then successfully negotiate and stipulate the pacts needed to establish it, as well as the legal framework to implement it. Transaction costs would need to be monitored closely and, whenever possible, reduced throughout the development phases.

Once the research universities became wholeheartedly committed to the idea of a regime that guaranteed them universal access to, and shared use of,

625. See Eben Moglen, *Why the FSF Gets Copyright Assignments from Contributors*, at <http://www.gnu.org/licenses/why-assign.html> (last visited Aug. 10, 2002); see also *supra* notes 572-581 and accompanying text.

626. See *supra* notes 599-601, 604-606 and accompanying text.

the government-funded data that they had collectively generated, these organizational problems might seem relatively minor. The difficulties of winning such a commitment, however, cannot be over-estimated in a world where university administrators are already conflicted about the efforts of their technology transfer offices to exploit commercially valuable databases in the genomic sciences and other disciplines with significant potential for commercial development. The prospect that Congress will eventually adopt a hybrid intellectual property right in collections of data could make these same administrators reluctant to lock their institutions into a kind of voluntary pool of any resulting exclusive property rights, even for public scientific research purposes.⁶²⁷

Conceptually, the problems inherent in organizing a pool of intellectual property rights so as to preserve access to, and use of, a common resource have become much better understood than in the past—owing to the experience gained from both the open-source software movement and the new Creative Commons initiative.⁶²⁸ These projects demonstrate that there are few, if any, technical obstacles that cannot be overcome by adroitly directing relevant exclusive rights and standard-form contracts to public, rather than private, purposes.

The deeper problem is persuading university administrators that they stand to gain more from open access to each others' databases in a horizontally organized research commons than they stand to lose from licensing data to each other under more restrictive, case-by-case transactions. The more that “the nature of the rivalry between . . . [universities] would shift from cooperative competition to turf wars, with rival networks of partners looking to delay, deter, and defend themselves against competitors,”⁶²⁹ the more they could make research data artificially scarce for them all. While we believe they stand to gain more from open access, following the implications of that conviction could amount to an act of faith, albeit one that resonates with the established norms of science and the primary mission of universities.

To the extent that universities may have to be sold on the benefits of an e-commons for data, with a view to rationalizing and modifying their disparate licensing policies, this project would require statesmanship, especially on the part of the leading research universities. It may also require pressure from the major government funders and standard-setting initiatives by scientific sub-communities. Funding agencies, in particular, must be prepared to discipline would-be holdouts and to discourage other forms of deviant strategic behavior that could undermine the cohesiveness of those institutions willing to pool their resources.⁶³⁰

627. Cf. Rai, *supra* note 33, at 113 n.204 (noting failure to negotiate an improved version of the Uniform Biological Materials Transfer Agreement of 1995 and major deviations from that Agreement in inter-university transactions).

628. See *supra* notes 571-579 and accompanying text.

629. Powell, *supra* note 13, at 266.

630. Account will have to be taken as well of the universities' patenting interests, which will need to be suitably accommodated.

Assuming a sufficient degree of organizational momentum, there remains the thorny problem of establishing the terms and conditions under which participating universities could contribute their data to a horizontally organized research commons. The bulk of the departments and sub-disciplines involved would almost certainly prefer a bright-line rule that required all deposits of government-funded data to be made without conditions and subject to no restrictions on use. This preference follows from the fact that most science departments currently see no realistic prospects for licensing basic research data, even to the private sector, and have not yet experienced the proprietary temptations of exclusive ownership that a *sui generis* intellectual property right in non-copyrightable databases might eventually confer.

At the same time, such a bright-line rule could utterly deter those sub-disciplines that already license data on commercial terms to either the private or public sectors, or that contemplate doing so in the near future. These sub-disciplines would not readily forego these opportunities and would, on the contrary, insist that any multilateral negotiations to establish a horizontal commons devise contractual templates that protected their commercial interests in the vertical dimension. If, moreover, Congress enacts a de facto exclusive property right in collections of data, it would probably deter other components of the scientific community, who might become unwilling to forego either the prospective commercial opportunities or other strategic advantages such rights might make possible.

A bright-line rule requiring unconditional deposits in all cases could thus defeat the goal of linking all university generators of government-funded data in a single, horizontally organized research commons. At the same time, the goal of universality could, paradoxically, require negotiators seeking to establish the system to deviate from the norm of full and open access by allowing a second type of conditional deposit of data into the horizontal domain by those disciplines or departments that are unwilling to jeopardize present or future commercial opportunities.

iii. Resolving the Paradox of Conditional Deposits

Science policy in the United States has long disfavored a two-tiered system for the distribution of government-funded data.⁶³¹ Under such a system, database proprietors envision split (or two-tier) uses of commercially valuable data and will only make them available on conditions that govern the different types of uses they have expressly permitted. In practice, there is growing evidence that, with regard to the exchange of biotechnology research tools, at least, university scientists “appear . . . to be creating a two-tiered market.”⁶³²

Formalized, split-level arrangements typically distinguish between relatively

631. Such two-tiered systems for government or academic data distribution have been favored and promoted by the scientific community in the E.U., but these initiatives have been strongly opposed by U.S. science agencies and academics. See, e.g., FULL AND OPEN EXCHANGE, *supra* note 8.

632. Eisenberg, *Bargaining*, *supra* note 13, at 242 (comparing the “free exchange tier” with the “proprietary tier” in the emerging two-tiered market in the exchange of research tools).

unrestricted uses for basic research purposes by nonprofit entities and more restricted uses for commercial applications by private firms that license data from scientific entities.⁶³³ The latter conditions may range from a simple menu of price-discriminated payment options to more complicated provisions that regulate certain data extractions, seek grant-backs of follow-on applications by second-comers, or impose reach-through clauses seeking legal or equitable rights in subsequent products.⁶³⁴ In some cases, moreover, the distinction between profit and nonprofit uses of scientific data becomes blurred, and the two categories may overlap, which adds to the cost and complications of administration.⁶³⁵ For example, universities may treat some databases as commercial research tools and impose a price discrimination policy that provides access to the research community at a lower cost than to for-profit entities.⁶³⁶

We recognize that a decision to allow participating universities to make conditional deposits of government-funded data to a collectively managed research consortium represents a second-best solution: one that conflicts with the goal of establishing a true public domain based on the premise of full and open access to all users. The allowance of restrictions on use breaks up the continuity of data flows across the public sector and necessitates administrative measures and transaction costs to monitor and enforce differentiated uses. It also entails measures to prevent unacceptable leakage between the horizontal and vertical planes, and may result in charges for public-interest uses that exceed the marginal cost of delivery, even in the horizontal plane.

We nonetheless doubt that a drive for totally unconditional deposits of government-funded data could succeed in the face of mounting worldwide pressure to commoditize scientific data,⁶³⁷ and we fear that excessive reliance on the

633. Cf. UNIF. BIOLOGICAL MATERIALS TRANSFER AGREEMENT (UBMTA) (1995), available at <http://www.autm.net/UBMTA/intro.html> (last visited Jan. 18, 2003) (allowing academic recipients of biological materials to use them, for noncommercial teaching and research purposes without having to negotiate a licensing agreement).

634. Cf., e.g., Eisenberg, *Patenting Research Tools*, *supra* note 123; Eckersley et al., *supra* note 610; Frischmann, *supra* note 70; Rai, *supra* note 34, at 113 n.204.

635. The importance of this distinction diminishes to the extent that any data provider—whether operating in industry or academia—actually accepts the conditions that contractually regulate the research commons. That is a major objective of our proposals.

636. This approach becomes likely when there is a private sector partner. Cf. Rai, *supra* note 34, at 111. Scientific data can also be made available in a “conditional public domain” through complicated three-way funding arrangements typically initiated by government science agencies under Cooperative Research and Development Agreements (“CRADAs”). Complications in this instance arise from tensions between the government’s continued interest in promoting public access and the legislative policies embodied in the Bayh-Dole Act, which encourage commoditization of government-funded research results. Even here, however, the fact that the government’s financial contribution to the project may predominate gives it the clout to impose conditions favorable to public-interest research uses. At present, this power is not sufficiently utilized. Cf. Bar-Shalom & Cook-Deegan, *supra* note 121; Rai & Eisenberg, *Bayh-Dole*, *supra* note 117, at 297-300. But a major purpose of establishing a solid legal framework for conditional deposits would be to provide standard-form licenses that clearly reinforce and implement favorable public-interest terms and conditions, without unduly compromising the relevant commercial interests.

637. European governments have already embarked on a policy of commercial exploitation of publicly generated data and even insist on conditional deposits in various governmental scientific organizations and cooperative research activities. Some academic scientific communities have recently tried to

orthodox position would, in the end, undermine—rather than save—the sharing ethos.⁶³⁸ Even if one disregards the prospects for strengthened intellectual property protection of non-copyrightable databases, too many universities have already begun to perceive the potential financial benefits they might reap from commercial exploitation of genomic databases in particular and biotech-related databases in general. Their reluctance to contribute such data to a research commons that allowed private firms freely to appropriate that same data could not easily be overcome.⁶³⁹

Even if a consortium of universities were to formally consent to such an unconditional arrangement, their technology transfer offices might soon be demanding an exceptional status for any databases that contained components produced without government funds.⁶⁴⁰ They could persuasively argue that private funds for most jointly created data products could decrease or even dry up if both customers and competitors could readily obtain the bulk of the data from the public domain. Once it became clear that an admixture of privately funded data could elicit the right to deposit data in a research commons on conditions that protected commercial exploitation of the databases in question, academics with an eye to cost recovery and profit maximization would logically make persistent efforts to qualify for this treatment. They would thus seek more private investment for this purpose or obtain the university's own funds for the project. Either way, there would be a perverse incentive to privatize more data than ever if the only legitimate way to avoid dedicating it all to the public domain was to show that some of it had been privatized.

In other words, if the quasi-public research space accommodated only unconditional deposits of data, it could foster an insuperable holdout problem as participating universities found ways to detach and isolate their commercially valuable databases from such a system. In these circumstances, a failure to obtain a best-case scenario premised on full and open access would quickly degenerate into a worst-case scenario, characterized by growing gaps in the communally accessible collection and an unraveling of the sharing ethos that

commercialize biotechnology databases of considerable public research value on a two-tiered basis, *see, e.g.*, Maurer, *Across Two Worlds*, *supra* note 34, while others have succeeded with controversial results, *see* the discussion of the Swiss-PROT at *supra* notes 599-606 and accompanying text. The reality is that U.S. universities intend to commercialize some of their data and support minimalist legislation to this end.

638. While we sympathize with the philosophy behind this position, our six years of focused study on issues concerning the legal protection of databases compels us to consider the realities of a growing trend toward two-tiered distributive activities to determine whether such activities can be operated in a manner that preserves the benefits of a public domain, notwithstanding the mounting pressures for commoditization. *See* Reichman & Samuelson, *supra* note 18; Reichman & Uhler, *Database Protection*, *supra* note 4.

639. Adoption of a database protection law would then magnify this reluctance and encourage the respective technology transfer offices to find more ways to commercially exploit more of the government-funded data products that were subsequently invested with proprietary rights. If the Bayh-Dole philosophy is factored into this equation, the prospects for persuading the universities both to agree and actually enforce a true public-domain model for all government-funded databases appear dim indeed. *Cf.* Rai, *supra* note 33, at 109-11 (describing impact of Bayh-Dole on prior research norms).

640. *Cf. id.* at 110-11 (describing growth of joint ventures and impact on MTAs).

would require case-by-case construction of inter-university data flows, and could sometimes culminate in bargaining to impasse.⁶⁴¹

In our estimation, the worst-case scenario is so bad, and the pressure to commoditize could become so great in the presence of a strong database right, that steps must be taken to ensure universal participation in a contractually reconstructed research commons from the outset by judiciously allowing conditional deposits of government-funded data on standard terms and conditions to which all stakeholders have previously agreed. Indeed, the goal is to develop negotiated contractual templates that clearly reinforce and implement terms and conditions favorable to public research without unduly compromising the ability of the consortium's member universities to undertake commercial relations with the private sector.

At stake in this process is not just a few thousand patentable inventions, but, rather, every government-funded data product that has potential commercial value to other universities as a research tool or educational device. Sound data management policies thus point to a second- best solution that would preserve the integrity of the inter-university commons by disallowing the principal ground on which concerted holdout actions might take root, by ensuring that only research-friendly terms and conditions applied in both the horizontal and vertical dimensions,⁶⁴² and by making it too costly for any institution to deviate from the agreed regulatory framework governing the two-tiered regime.

Those who object to this proposal will argue that it unduly undermines the full and open access principle by tempting more and more university departments or sub-disciplines to opt for conditional deposits than would otherwise have been the case. On this view, once a negotiated two-tiered model was set in place, universities would come under intense pressure to avoid the true public domain or open access option even when there was no need to do so.

However, a universal and functionally effective inter-university research commons simply cannot be constructed with a bright-line, true public domain rule applied across the board for the reasons we have previously set out. A bright-line rule also carries with it the well-recognized difficulty of distinguishing for-profit from not-for-profit research activities when single laboratories increasingly engage in both. In contrast, a regime based on conditional deposits overcomes this problem by allowing a scientific entity to contribute to and benefit from the data commons so long as it respects the agreed norms bearing on that arrangement. In this respect, a normative accommodation will have displaced legal distinctions that cannot feasibly be enforced.⁶⁴³

641. See, e.g., Powell, *supra* note 13, at 266 (predicting “disputes, duplication, and discord” in analogous situations).

642. Data managed by a consortium would presumably also be subject to its agreed contractual templates regulating the licensing of government-funded data to the private sector. See discussion *supra*, at Part IV.C.1.c..

643. Cf. Rai, *supra* note 33 (expressing preference for normative solutions).

Moreover, the very contractual templates that make the construction of such a commons feasible in a two-tiered system should also mitigate its social costs. Even if conditional deposits are allowed, many sub-disciplines will continue to have no commercial prospects and no need to invoke the contractual templates that regulate them. When this is the case, peer pressure reinforced by the funding agencies should make it difficult, if not impractical, for members of those communities to opt out of the traditional practice of making data available unconditionally.⁶⁴⁴

When, instead, given communities find themselves forced to deal with serious commercial pressures, the negotiated contractual solutions that enable them to make data conditionally available for public research purposes should also tend to preserve and implement the norms of science. In particular, the applicable contractual templates should immunize deposited data from the vagaries of case-by-case transactions under the aegis of university technology transfer offices⁶⁴⁵ and should also limit the kinds of restrictions private-sector partners might otherwise seek to impose on universities.

At the end of the day, a set of agreed contractual templates permitting conditional deposits in the interest of a horizontally linked research commons would provide a tool universities could use with more or less wisdom. If used wisely, this tool should ensure that more data are made available to a contractually reconstructed research commons than would be possible if member universities could not protect the interests of their commercial partners. This same tool may also provide incentives for the private sector to work with universities to produce better data products than the latter alone could generate with their limited funds.

iv. Other Hard Problems

Allowing universities to deposit government-funded data into a contractually reconstructed research commons, on conditions designed to protect their commercial relations with the private sector, solves two difficult problems. First, it avoids the risk that large quantities of government-funded data would remain outside the system on the ground that they had been commingled with privately funded components. Second, it ensures that any negotiated contractual templates the research consortium adopts to govern its horizontal space will apply to all the data holdings within its jurisdiction, including databases to which the private sector had contributed. However, it does not automatically determine the precise conditions that the agreed contractual templates should apply to inter-university licensing of data within their collective jurisdiction. In the process of defining these conditions, moreover, those who negotiated the

644. *Cf. id.* at 112 (“Major research universities have sought to maintain certain aspects of traditional scientific norms even while embracing the development-promoting aspects of property rights.”) (citing examples).

645. *See, e.g.,* Eisenberg, *Bargaining*, *supra* note 13, at 243 (stressing the extent to which technology transfer professionals “become ever more wary of free exchange and more assiduous about restricting its domain”).

multilateral pact between universities, federal funders, and scientific bodies needed to launch the consortium would have to resolve a number of contentious issues.

The guiding principle that should apply to inter-university licensing of data available from the quasi-public space is that depositors may not impose any conditions that impede the customary and traditional uses of scientific data for nonprofit research purposes. A logical corollary is that they should affirmatively adopt the measures that may prove necessary to extend and apply this principle to the online environment.⁶⁴⁶ Because the data under discussion are government-funded for academic purposes to begin with, the open access and sharing norms of science should then color any specific implementing templates that regulate access and use.

(a) *Access.* With regard to access, the customary mode of implementing these norms would be to make data available to other nonprofit institutions at no more than the marginal cost of delivery. In the online environment, these marginal costs are essentially zero. This represents the preferred option whenever the costs of maintaining the data collection are defrayed by public subsidy or by non-exclusive licenses to private firms in the vertical dimension.

If, however, the policy of free or marginally priced access appears unable to sustain the cost of managing a given project at the inter-university level, an incremental pricing structure may become unavoidable. The options for such a pricing structure range from a formula allowing partial incremental cost recovery when a project is partially subsidized to a formula providing full cost recovery when this is necessary to keep the data collection alive.⁶⁴⁷ Examples of sub-communities that have found it necessary to rely on the second option are largely in the laboratory physical sciences.⁶⁴⁸

The prices charged other nonprofit users to access data in the research commons should never exceed the full incremental cost of managing the collective holdings. This premise follows from the fact that the initial cost of collecting or creating the data was defrayed by the government or some combination of sources (including private sources) that normally subscribe to the open access principle.

When, however, private firms have defrayed a substantial part of the cost of generating the database in question, there are few, if any, standard solutions.

646. Cf. WIPO Copyright Treaty, *supra* note 268, at art. 10; Agreed Statement Concerning art. 10, WIPO.doc.CRNR/DC/96 (Dec. 20, 1996).

647. This may be accomplished through a paying inter-university consortium, such as the Inter-University Consortium for Political and Social Research (“ICPSR”) at the University of Michigan, or by means of more ad hoc cost-recovery methods. Institutional membership dues for the year 2002 ranged from \$2,000 per year for institutions in developing countries to \$15,000 per year for full membership status by large U.S. institutions. See Inter-University Consortium for Political and Social Research, <http://www.icpsr.umich.edu/> (last visited Jan. 18, 2003).

648. Databases operated on a cost recovery basis may be found primarily in the areas of ceramic, pharmaceutical, chemical, and metallurgical research, as well as in state-of-the-art manufacturing operations. Telephone Interview with Prof. Robert L. Snyder, Ohio St. Univ. (Aug. 30, 2002).

Occasionally, even a private partner might view the collective holdings as a valuable resource for its own pursuits, to which it agrees to contribute on an eleemosynary basis.⁶⁴⁹ In the more typical cases, the private partner is likely to view the research community as the target market for a database it paid to create and from which it must derive its expected profits.⁶⁵⁰

In that event, the collection of additional revenues from private sector access charges should depend entirely on freedom of contract, while a likely demand that public research users pay access charges that exceed data management costs would pose a hard question. On the one hand, as beneficiaries of government funding, the universities should forego profits from charges levied to access their partly publicly funded databases for public research purposes. On the other hand, a private partner will not readily forego such profits, especially if it had invested in the project precisely because of its potential commercial value as a research tool.⁶⁵¹ If the university shared these profits with its private partner, this practice would deviate from the basic principle governing inter-university access generally, and would encourage other universities to seek private partners for this purpose, which in turn would yield both social costs and benefits.

In these cases, care must be taken to avoid adopting policies that would discourage either the formation of public-private partnerships for the development of socially beneficial data products or the inclusion of such products in a horizontal, quasi-public research space. At the same time, there is a potential loophole here that would allow universities to deviate from the general rules applicable to that space if the private partner could impose market-driven access rights for nonprofit research purposes, and its partner university shared in those profits.

We know of no standard formula for resolving this problem. If the database is also of interest to the private sector, price discrimination and product differentiation are the preferred techniques for reducing access charges levied for public research. In any event, the trustees that manage the inter-university system should monitor and evaluate these charges, and their power to challenge unreasonable or excessive demands would become especially important in the absence of any alternative or competing source of supply.⁶⁵²

This strategy, however, begs the question of whether and to what extent the universities should be allowed to retain their share of the profits from access

649. Cf. Rai, *supra* note 33, at 114 (discussing case of Celera's refraining from claiming certain intellectual property rights in the genome).

650. Cf. *id.* at 111, 141 (documenting restrictive conditions imposed upon researchers under MTAs arising from joint ventures between universities and private firms).

651. See, e.g., Eisenberg, *Bargaining*, *supra* note 13, at 229 ("When one research institution's research tool is another firm's end product, it is difficult to agree upon a universe of materials that should be exchanged on standardized terms.").

652. Because the data in question are partly government-funded by definition, a reasonable terms and conditions clause should automatically apply.

charges levied against public research users.⁶⁵³ As matters stand, this is an issue that can only be addressed by the relevant discipline communities themselves, in the absence of some general norm that would not pose insuperable administrative burdens to implement.

(b) *Use.* Once access to databases available to the research commons has legitimately been gained, further restrictions on uses of the relevant data should be kept to a minimum. In principle, contractual restrictions on reuse of publicly funded data for nonprofit research purposes should not be permitted. This principle need not impede the use of conditions that require attribution or credit from researchers who make use of such data,⁶⁵⁴ and it can also be reconciled with provisions that defer access by certain users for specified periods of time or impose restrictions on competing publications for a certain period of time.⁶⁵⁵

This ideal principle runs into trouble, however, when confronted with the difficult problem posed by commercially valuable follow-on applications derived from databases made available to the research commons. It is one thing to posit that the academic beneficiaries of publicly funded research should be limited to the recovery of costs through access charges and should not be entitled to additional claims for follow-on uses by other nonprofit researchers. Quite different situations arise when the funding is public, but a private firm has invested its own resources to develop a follow-on application for commercial pursuits or when the initial data-generating project entailed a mix of private and public funds and the product subsequently gives rise to a commercially valuable follow-on application. These hard cases become even harder if the follow-on product primarily derives its commercial value from being a research tool universities themselves need to acquire.

Assuming, as we do, that a primary objective of any negotiated solution is to avoid gaps in the data made available for public research purposes in the horizontal domain, there is an obvious need for agreed contractual templates that would respect and preserve the commercial interests in the vertical plane identified above. This goal directly conflicts, however, with the most idealistic option set out above, which is to freely allow all follow-on applications based on data made available to the research commons, regardless of the commercial prospects or purposes and without any compensatory obligation beyond access charges (if any).

This option would represent a true public domain approach to government-funded data and would fit within the traditional legal framework applied in the past to collections of data. However, it might be expected to discourage public-private partnerships formed to exploit follow-on applications of publicly funded

653. If these charges were administered by trustees or designated agents, they could be redistributed in the form of grants. *Cf.* Eckersley et al., *supra* note 610, at 10.

654. *See, e.g.*, discussion of the Creative Commons templates, *supra* note 579.

655. *Cf.* Patrinos & Drell, *supra* note 23.

databases, contrary to the philosophy behind the Bayh-Dole Act, although this risk is tempered by the fact that all would-be competitors who invested in such follow-on applications would find themselves on equal footing in this respect. This option would certainly discourage public-private partnerships formed to produce scientific databases from making them available to the commons if that decision automatically deprived them of any rights to follow-on applications.

A second option is to leave the problem of commercially valuable follow-on applications to freedom of contract, in which case universities and their private partners could license whom they please and exclude the rest. This solution is consistent with proposals to enact a de facto exclusive property right in non-copyrightable databases and with the philosophy behind Bayh-Dole. It would also alleviate disincentives to make databases derived from a mix of public and private funds available to the nonprofit research community.

However, this second option would relegate the problem of follow-on applications to the universities' technology transfer offices once again, which might be tempted routinely to impose the kind of grant-back and reach-through clauses that are already said to generate anticommons effects in biotechnology⁶⁵⁶ and are inconsistent with the dual nature of data as both inputs and outputs of innovation. Just as a true public domain approach tends unintentionally to impoverish the commons we seek to construct, so too a true laissez-faire approach undermines the effectiveness of that same commons and triggers a race to the bottom, as universities seek private partners solely for the purpose of occupying a privileged position with respect to follow-on applications.

A third option is freely to allow follow-on applications of databases made available to the research commons for commercial purposes while requiring their producers to pay reasonable compensation for such uses under a predetermined menu that fixes a range of royalties for a specified period of time.⁶⁵⁷ For maximum effect, a corollary "no holdout" provision should obligate all universities engaged in public-private database initiatives to make the resulting databases available to the research commons under this compensatory liability framework.

This approach would enable investors in public-private database initiatives to make their data available for public research purposes without depriving them of revenue flows from follow-on applications needed to cover the costs of R & D or the opportunities to turn a profit. At the same time, it would avoid impeding access to the data for either commercial or noncommercial purposes, in which aspect it would mimic a true public domain and create no barriers to entry.⁶⁵⁸ Moreover, a compensatory liability approach would implement the

656. See, e.g., Rai, *supra* note 33 at 149.

657. For details, see discussion of compensatory liability approach in Reichman, *Of Green Tulips*, *supra* note 164. See also Reichman, *Legal Hybrids*, *supra* note 164; Reichman & Samuelson, *supra* note 18, at 145-51.

658. Transaction costs can be kept low by means of standard-form deals administered by the managers of the system. See Eckersley et al., *supra* note 610, at 10-11 (proposing collective licensing organization for analogous purposes).

policies behind the Bayh-Dole Act without the overkill that occurs when publicly funded research results are subjected to exclusive property rights that impoverish the public domain and create barriers to entry to boot.

These, or other, options would require further study and analysis as part of the larger process of reconstructing the research commons we propose. It should be clear, moreover, that any solutions adopted at the outset must be viewed as experimental and subject to review in light of actual results.

2. Informal Data Exchanges

As constituted at the present time, the zone of informal data exchange is populated by single researchers or laboratories or by small teams of associated researchers whose work is typically expected to lead to future publications. Because this zone operates largely in a pre-publication environment, the constraints of government funders on uses of data are relatively less prescriptive, and a considerable amount of the data being produced may not be funded by federal agencies at all. If funding is provided by other nonprofit sources or by state governments, the end results still pertain to public science and its ultimate disclosure norms, but the controls are not standardized. To the extent that private sector funding is also involved, even the norms of public science may not apply.

Quantitatively, the amount of scientific data held in this informal zone appears large. Despite the relative degree of invisibility that pre-publication status confers, these holdings are also of immense qualitative importance for cutting-edge research endeavors. Although these data may not be as well prepared as those released for broad, open use in conjunction with a publication, they typically reflect the most recent findings.

Moreover, this informal sector seems destined to grow even more important in the near future as it increasingly absorbs scientific data that were not released at publication as well as data researchers continue to compile after publication. If Congress were to adopt a strong intellectual property right in non-copy-rightable databases, this informal zone could expand further to include all the published data covered by an exclusive property right that had not otherwise been dedicated to the public domain.

As previously discussed, actual secrecy is taken for granted in this zone, and disclosure depends on individually brokered transactions often based on reciprocity or some *quid pro quo*.⁶⁵⁹ These fragile data streams, which have always been tenuous due to personal and strategic considerations, have increasingly broken down owing to denials of access and to a trading mentality steeped in commercial concerns that is displacing the sharing ethos.

Our previous analysis showed that, left to themselves, the legal and economic pressures operating in the informal zone are likely to further reduce disclosures over time and make the informal data exchange process resemble that

659. See *supra* notes 482-484 and accompanying text.

of the private sector.⁶⁶⁰ That trend, in turn, undermines the new opportunities to link even highly distributed data holdings in virtual archives or to experiment with new forms of collaborative research on a distributed, autonomous basis, as digital networks have recently made possible. The positive synergies expected from organized peer-to-peer file sharing on an open access basis cannot be realized if researchers decline to make data available at all out of a fear of sacrificing new-found commercial opportunities or other strategic advantages. Nor will these new opportunities fully develop if those who are nominally willing to make data available impose onerous licensing terms and conditions—reinforced by intellectual property rights—that multiply transaction costs, unduly restrict the range of scientific uses permitted, or otherwise embroil those uses in anti-commons effects.

Here, the immediate goal of science policy should be to reduce the technical, legal, and institutional obstacles that impede electronic peer-to-peer file exchange and to generally facilitate exchanges of data on the most open terms possible across a horizontal or quasi-public space. At the same time, the measures adopted to implement this policy must avoid compromising or inhibiting the interests of individual participants who seek commercial application of their research results in a private or vertical sphere of operations. This two-pronged approach could stabilize the status quo and reinvigorate the flagging cooperative ethos in the zone of informal data exchange, as more individual researchers and small communities experience the benefits of electronically linked access to virtual archives and discover the productive gains likely to flow from collaborative, interdisciplinary, and cross-sectoral uses.

From an institutional perspective, however, organizing and implementing such a two-pronged approach to data exchange in the informal zone presents certain difficulties not encountered in the formal zone of inter-university relations. Here, the playing field is much broader, the players are more autonomous and unruly, and the power of federal funders directly to impose top-down regulations has traditionally been weak or under-utilized. The moral authority of these funders nonetheless remains strong, and peer pressure in support of the sharing ethos would become more effective if a consensus developed that the two-pronged approach we envision actually yields tangible benefits at acceptable costs.

Much therefore depends on short-term, bottom-up initiatives that rely on individual decisions to opt for standardized, research-friendly licensing agreements in place of the defensive, ad hoc transactions that currently hinder the flow of data streams in this sector. The solution is to provide individual researchers with a toolkit for constructing prefabricated exchange transactions on community-approved terms and conditions. The toolkit would contain a menu of standard-form contractual templates that individual researchers could use to license data, and the templates adopted would be posted online to facilitate

660. See *supra* Part III.C.1.a.ii (discussing the effects of a highly protectionist regime on informal data exchanges).

electronic access to networks of nodes.⁶⁶¹ These templates would cover a variety of situations and offer a range of ad hoc choices, all aimed at maximizing disclosure in both digital and non-digital mediums for public research purposes.

For this endeavor to succeed, however, the templates in question would clearly need to allow participating researchers and their communities to make data available on conditions that expressly preclude licensees from unauthorized commercial uses or follow-on applications. While this suggests the need to deviate from true public domain principles once again, one should remember that, in the informal zone as it stands today and is likely to develop, secrecy and denial of access are already well-established, countervailing practices. One can hardly argue that permitting conditional availability would undermine the norms of science in this zone, given the inability of those norms to adequately defend the interests of public research in unrestricted flows of data at the present time.

The object is, rather, to invigorate those sharing norms by reconciling them with the commercial needs and opportunities of the researchers operating in the informal zone, in order to elicit more overall benefits for public science under a second-best arrangement than could be expected to emerge from brokered individual transactions in a high-protectionist legal environment. This strategy requires a judicious resort to conditionality that would make it possible to forge digitally networked links between individual data suppliers and that would let their data flow across those links into a quasi-public space relatively free of restrictions on access and use for commercial purposes.

Given the large number of players and the disparity of interests at stake, a logical starting premise is that only a small number of standard contractual templates seems likely to win the support of the general scientific community, at least initially. A true public domain option should, of course, be available for all willing to use it. For the rest, a limited menu of conditional public domain provisions, such as those offered by the Creative Commons, should suffice.⁶⁶² Clauses that delay certain uses for a specified period, or that delay competing publications based on, or derived from, a particular database for a specified period of time should also pass muster, so long as they remain consistent with the practices of the relevant scientific sub-community. In the absence of any underlying intellectual property right, an additional clause reserving all other rights and excluding unauthorized commercial uses and applications would complete the limited, “copyleft” concept. We believe that even a small number of standard contractual templates that facilitate access to and use of scientific data for public research purposes could exert a disproportionately large impact on the increasingly open, collaborative work in the networked environment.

In the scientific milieu, however, difficult problems of leakage and enforcement could also arise. To address these problems, the scientific community,

661. See Creative Commons, *supra* note 571 and accompanying text.

662. See discussion *supra*, note 579.

perhaps under the auspices of the American Association for the Advancement of Science, would need to consider developing institutional machinery capable of assisting individual researchers who feared that their data had been used in ways that violated the terms and conditions of the standard-form licensing agreements they elected to employ.

More complex or refined contractual templates are also feasible, but their use should normally depend less on individual choice and more on the consensus approval of discipline-specific communities. Moreover, in the informal zone, efforts to influence the terms and conditions applicable to private-sector uses seem much less likely to succeed than similar efforts in the inter-university context.⁶⁶³

Attempts to over-regulate the zone of informal data exchange should generally be avoided at this stage, lest they stir up unwarranted controversy and deter the more ambitious efforts to regulate inter-university transactions described above. The success of those efforts in the zone of formal data exchanges should greatly reinforce the norms of science generally. It would also exert considerable indirect pressure on those operating in the informal zone to respect those norms and emulate at least the spirit of any agreed contractual templates that had proved their merit in that context.

The more that universities succeed in amalgamating their government-funded holdings into an effective, virtual archive or repository, the more pressure that would bring to bear on individual researchers, research teams, and small communities to similarly make their data available in more formally constituted repositories. As a body of practice develops in both the formal and informal zones, the most successful approaches and standards will become broadly adopted, and the desire to obtain the greater benefits likely to flow from more formalized arrangements should grow.⁶⁶⁴

Meanwhile, efforts to regulate the zone of informal data exchanges should be viewed as an opportunity to strengthen the norms of science and to facilitate the creation of virtual networked archives electronically linking disparate and highly distributed data-holders. The overall objective should be to generate more disclosure than would otherwise have been possible if all the players exercised their proprietary rights in total disregard of the need for a functioning research commons for nonprofit scientific pursuits. If successful, these modest efforts in the informal zone could alleviate some of the most disturbing erosions of the sharing ethos that have already occurred, and could encourage federal funding agencies to take a more active role in regulating broader uses of research data. A successful application of “copyleft” techniques to the informal zone of academic research could also serve as a model for encouraging disclosure for public research purposes of more data generated in the private sector.

663. It seems unlikely, for example, that any standard grant-back or reach-through clauses for commercial applications of databases could serve the varied interests of the different communities operating in the informal zone.

664. *See supra* note 133 and accompanying text.

D. Proposals for the Private Sector

Scientific data produced by the private sector are logically subject to any and all of the proprietary rights that may become available, as surveyed earlier in this article.⁶⁶⁵ Here, the policy behind a contractually reconstructed research commons is not to defend the norms of science so much as to persuade the private sector of the benefits it stands to gain from sharing its own data with the scientific community for public research purposes. The goal is thus to promote voluntary contributions that might not otherwise be made to the true public domain or to the conditional domain for public research purposes on favorable terms and conditions.⁶⁶⁶

From the perspective of public-interest research, of course, corporate contributions of otherwise proprietary data to a true public domain is the preferred option. While the copyright paradigm reflected in the Supreme Court's *Feist* decision presumably made the factual contents of commercially valuable compilations published in hard copy available for such purposes,⁶⁶⁷ some federal appellate courts have lately rebelled against *Feist* and made it harder for second-comers to separate non-copyrightable facts and information from the elements of original selection and arrangement that still attract copyright protection.⁶⁶⁸

Online access to non-copyrightable facts and data is further restricted by the stronger regime that prohibits tampering with technological fences that was embodied in the DMCA,⁶⁶⁹ although the full impact of these provisions on scientific pursuits remains to be seen. Meanwhile, many commercial database publishers may be expected to continue to lobby hard for a strong database protection law on the E.U. model that would limit unauthorized extraction or reuse of the non-copyrightable contents of factual compilations, and it appears likely that Congress will seek to enact a database protection statute in 2004.⁶⁷⁰

In contrast to the research-friendly legal rules under the print paradigm, all the factual data and non-copyrightable information collected in proprietary databases are increasingly unlikely to enter the public domain and will instead come freighted with the restricted licensing agreements, digital rights management technologies, and *sui generis* intellectual property rights that characterize a high-protectionist legal environment. Under such a regime, open access and unrestricted use become possible only if private-sector database compilers donate their data to public repositories or contractually agree to waive proprie-

665. See discussion *supra*, Parts II.C, III.B.2.

666. Cf. Rebecca S. Eisenberg, *Intellectual Property at the Public-Private Divide: The Case of Large-Scale cDNA Sequencing*, 3 U. CHI. L. SCH. ROUNDTABLE 557, 559 (1996); Rai, *supra* note 33, at 134 (discussing Merck & Co.'s practice of putting into the public domain the results of an EST identification project it sponsored at Washington University).

667. See *supra* notes 78, 85-89 and accompanying text.

668. See *supra* notes 283-287 and accompanying text.

669. See discussion *supra*, Part III.B.2.b.

670. See *supra* note 376.

tary restrictions on controls that would otherwise impede access and use for public research purposes.⁶⁷¹

Some examples of both donated and contractually stipulated public domain data collections from the private sector already exist. In the first category, for instance, three major energy companies recently donated large collections of proprietary rock samples to the University of Texas at Austin.⁶⁷² The samples and related data are managed as a public research resource by the University. The companies also donated land, buildings, equipment, and additional funds to provide the physical infrastructure and a partial endowment for operating expenses, while retaining no proprietary interest in the donated materials.⁶⁷³

An example of the second type of arrangement is the SNP Consortium Ltd., a nonprofit foundation created by thirteen pharmaceutical and information technology companies and Wellcome Trust, whose stated purpose is to provide genomic data to the public domain.⁶⁷⁴ The motivation for this apparent corporate largess was to save substantial money by pooling resources and to prevent any other private sector entity from capturing the data or otherwise encumbering access to them.⁶⁷⁵ The Consortium partners have found this approach to be very successful. Instead of spending \$250 million to identify 150,000 single nucleotide polymorphisms (“SNPs”),⁶⁷⁶ as was originally estimated by one of the partners, the shared project cost amounted to \$44 million and yielded almost 1.8 million SNPs—all freely available to researchers on the open website.⁶⁷⁷ Two other projects are now being planned by the SNP Consortium partners using the same public domain model—a protein structures consortium and a public DNA database in the United Kingdom, the latter in collaboration with the Department of Health and the Medical Research Council.⁶⁷⁸

Although pure public domain models initiated by industry will no doubt continue to be the exception rather than the rule, the availability of data on a conditional public domain basis, or at least on preferential terms and conditions to the not-for-profit research community, should enjoy far broader acceptance and ought to be promoted. Certainly, the existence of contractual templates,

671. This will also depend to a large extent on the exceptions for nonprofit research built into the database protection law itself.

672. See Shirley Dutton, *Corporate Donations of Geophysical Data*, in NRC SYMPOSIUM, *supra* note 10; see also NATIONAL RESEARCH COUNCIL, *GEOSCIENCE COLLECTIONS AND DATA: NATIONAL RESOURCES IN PERIL*, (2002) at 25 (detailing no examples of transfer from corporate-owned repositories to state geological surveys).

673. Dutton, *supra* note 672.

674. See SNP Consortium at <http://snp.cshl.org/> (last updated Feb. 5, 2003).

675. Michael Morgan, *The SNP Consortium*, in NRC SYMPOSIUM, *supra* note 10.

676. Single nucleotide polymorphisms are common DNA sequence variations among individuals. Although approximately 99.9% of the three billion nucleotides in each person are identical, there are nonetheless three million differences in the remaining 0.1% that are distributed throughout the three billion nucleotides. By identifying and mapping out these differences, it then becomes possible to correlate them with human susceptibility to disease (for example, diabetes or cancer) and their responsiveness to various drug therapies. See Morgan, *supra* note 675.

677. *Id.*

678. *Id.*

along the lines being developed by the Creative Commons, could help to encourage private sector entities to make conditional deposits of data for relatively unrestricted access and use by public-interest researchers.⁶⁷⁹ Indeed, some enlightened CEOs have acknowledged the benefits that derive from enriching the contents of an information commons that all researchers can use for further innovation,⁶⁸⁰ and worldwide efforts to compile databases of traditional knowledge gleaned from indigenous populations point in the same direction.⁶⁸¹

Scientific publications by private sector scientists provide another valuable source of research data. However, these scientists labor under increasing pressures either to limit such publications altogether or to insist that publishers allow supporting data to be made available only on conditions that aim to preserve their commercial value.⁶⁸² Although many academics in the scientific community oppose this practice,⁶⁸³ it is exactly what would proliferate if private sector scientists held exclusive property rights in the data that allowed them to retain control even after publication. This sobering observation might induce the scientific community to reconsider the need to allow private sector scientists to modify the bright-line disclosure rules otherwise applied to public sector scientists, in order to encourage them to disclose more of their data for nonprofit research purposes.

Even when companies remain unwilling to make their data available to nonprofit researchers on a conditional public domain basis, there is ample experience with price discrimination and product differentiation measures favorable to academics.⁶⁸⁴ To the extent that the public research community does not constitute the primary market segment of the commercial data producer, either of these approaches would help promote access and use by noncommercial researchers without undue risks to the data vendor's bottom line. The conditions under which such arrangements might be considered acceptable by commercial data producers will vary according to discipline area and type of data product, but it is in the interest of the public research community to identify such producers in each discipline and sub-discipline and to negotiate favorable access and use agreements on a mutually acceptable basis.

The terms and conditions acceptable to private firms that do opt to deposit data into a public access commons arrangement might be fairly restrictive in their allowable uses, as compared with the conditions applicable under the

679. *See supra* note 579.

680. *See* Edward Shonsey, *Biotechnology: An Idea Before Its Time?*, Presentation at The University of Minnesota Conference on Governing GMOs: Developing Policy in the Face of Scientific & Public Debate, (Feb. 1, 2001), *available at* <http://www.lifesci.consortium.umn.edu/conferences/gmosconf/> (last visited Jan. 18, 2003).

681. *See, e.g.*, Dutfield, *supra* note 114; *Traditional Knowledge, Intellectual Property and Indigenous Culture*, Symposium presented at the Benjamin N. Cardozo School of Law (Feb. 21-22, 2002). *But see* van Caenegem, *supra* note 113, at 328-30 (stressing risks of dispossessing indigenous claims to knowledge).

682. *See generally* COMMUNITY STANDARDS, *supra* note 74.

683. *Id.* (views of the biological sciences community).

684. *See, e.g.*, BITS OF POWER, *supra* note 1, at 124-25.

standard-form templates that researchers themselves would normally adopt. Nevertheless, the goal of securing greater access to privately generated data with fewer restrictions justifies this approach because it would make data available to the research community that would otherwise be subject to commercial terms and conditions in a less research-friendly environment.

Finally, the importance of regulating the interface between university-generated data and private sector applications was treated at length above, with a view to ensuring that the universities' eagerness to participate in commercial endeavors did not compromise access to, and use of, federally funded data for public research purposes.⁶⁸⁵ Here, in contrast, it is worth stressing the benefits that can accrue from data transfers to the private sector whenever a framework for reducing the social costs of such transfers has been worked out to the satisfaction of both the research universities and the public funding sources. These arrangements are especially important if the exploitation, or applications of, any given database by the private sector would not otherwise occur in a non-proprietary environment.

Price discrimination and product differentiation can also facilitate socially beneficial interactions between the private sector and universities. For example, companies might consider licensing certain data to commercial customers on an exclusive-use basis for a limited period of time, after which the data in question would be licensed on preferential terms to nonprofit users or even revert to an open access status. This strategy might work successfully in the case of certain environmental data, in which most commercially valuable applications are produced in real time or near-real time and can then be made available at lower cost and with fewer restrictions for retrospective research that is less time dependent.⁶⁸⁶ Such an approach might not work in other research areas, such as biotechnology, however, in which a delay in access may not be an acceptable tradeoff or that delay is too long to preserve competitive research values.⁶⁸⁷

Especially serious problems seem likely to arise when the public research community becomes the target market for the commercial data supplier, and there is a resulting tension between freedom of contract and the needs and capabilities of the nonprofit research sector. In principle, one expects that a supplier will not price itself out of the market. In practice, some science publishers have adopted exorbitant pricing strategies that do limit scientists' abilities to access and use their products.

If a database protection law is enacted and sole-source science publishers control databases of major importance to scientific sub-communities, access and use will increasingly depend on any exceptions and immunities favoring

685. See discussion *supra*, Part IV.C.i.c.

686. See, for example, the Sea Wide Field Sensor example in the NASA data purchase program, at <http://seawifs.gsfc.nasa.gov/SEAWIFS.html> (last visited Jan. 18, 2003).

687. The trend in industry may in fact lie in the direction of less flexibility and accommodation to nonprofit research interests. See Nelson, *supra* note 132.

research that are built into the law, and on any misuse provisions requiring database licensors to adopt reasonable terms and conditions. If and when these problems become acute, as may well happen, the science community will have to consider appropriate actions including both collective bargaining arrangements (to the extent permitted by antitrust laws) and concerted efforts to independently develop alternative sources of supply for public research purposes.

V

CONCLUSION

The importance of public domain data for scientific research is so taken for granted that it becomes difficult to identify the precise boundaries of that domain, to describe its operations, and especially to evaluate the normative and legal infrastructure that supports those operations. At the outset of this article, we undertook to map the public domain for scientific data as it actually functions today, with a view to addressing two new challenges. One is the advent of digital networks, which is transforming the traditional modes of exchanging scientific data and could considerably magnify the payoffs that science (and industry) derive from a policy of full and open access to public research data. The other is an array of economic, legal, and technical pressures that threaten to impede or disrupt the continued operations of that same public domain for scientific data as it was traditionally constituted.

Our investigation reveals that the policy of open access to public research data rests on a surprisingly fragile foundation in both the legal and normative sense. As scientists and universities increasingly aspire to commercialize their research products (partly in response to the Bayh-Dole Act and related legislation), their willingness to exchange data, along with other research tools, has begun to suffer. There is evidence that informal exchanges of data in some fields, such as biomedical research, have become severely compromised, while inter-university exchanges are subject to high transaction costs, delay, and a growing risk of anticommons effects. As relations between universities and industry become more intense, the ability of the industrial partners to impose restrictions on the open availability of research data also increases and could pose a formidable obstacle in the future.

In this already delicate situation, the advent of strong new intellectual property rights in databases could have disproportionately adverse effects on the operations of the public domain for scientific data. A database protection law would remove data from their traditional public domain status under copyright law. It could invest scientists and universities with exclusive property rights in collections of data—including government-funded databases—that would survive both the publication of research results in scientific journals and the disclosure of such results in patent applications. Database rights, when added to other economic and technical pressures, could thus become the hub of an enclosure process that progressively fences off the public domain for scientific data and undermines its functions. This process could greatly reduce the flow of

data as a basic input into both scientific research and the national system of innovation.

We have argued that science policy should take steps now to address these challenges to ward off the threat of undue enclosure and to exploit of the potential benefits of digitally linked data resources. We focused particular attention on government-funded data because of its overall importance to the scientific enterprise and because it already benefits from a regulatory structure that could be appropriately adjusted and strengthened to preserve crucial public domain functions even in a highly protectionist intellectual property environment. We suggest that science policy should treat data produced with government funds as a collective resource for research purposes. Government agencies, research universities, and scientific bodies should accordingly negotiate and develop a regulatory framework to preserve the functions of a research commons by contractual means that constrain private rights to serve the public interest.

It is not necessary for this purpose that universities forego their growing opportunities to participate in commercial applications of research results. It is necessary, however, to curb the ability of their industrial partners to restrict the flow of data as a collective research resource. It is even more necessary to develop a strong legal and normative infrastructure that would preserve open access to, and use of, the data generated by participating universities for public research purposes. However, this outcome can only be achieved by realistic contractual arrangements that also preserve the participants' ability to license data to the private sector while ensuring that scientists can access and use the same data on reasonable terms and conditions.

We believe that science policy stands at a critical threshold. If nothing is done to address the challenges we identify, the unraveling of the sharing ethos that already characterizes what we have termed the zone of informal data exchanges between individual scientists will spread to universities, and a trading mentality will further contaminate inter-university exchanges of data.

If, instead, science policy takes timely action to address these problems, the benefits could be spectacular, given the new opportunities for scientific collaboration that digital networks make possible. If government-funded data at the university level do enter a contractually reconstructed research commons along the lines we advocate, it would put considerable pressure on single scientists and laboratories to conform their own data exchange practices to the broader normative and regulatory ethos by means of suitable contractual templates. The formulation of these templates could, in turn, make it possible to link up the highly distributed databases of cutting-edge disciplines into "networks of nodes." On this scenario, the research commons—instead of shrinking and becoming increasingly dysfunctional—could yield positive externalities and network effects that exceeded anything that the scientific community had previously experienced.