**PRESENTED AT THE CENTRE DE PHILOSOPHIE DU DROIT CONFERENCE
ON MICROBIOLOGICAL COMMONS,
UNIVERSITÉ CATHOLIQUE DE LOUVAIN,
BRUSSELS, JULY 7-8, 2005**

IDENTIFYING AND ADDRESSING GLOBAL TRENDS TO RESTRICT ACCESS TO SCIENTIFIC DATA FROM GOVERNMENT FUNDED RESEARCH

by Jerome Reichman, Duke University School of Law &
Paul Uhlir, The National Academies, U.S.A.

IDENTIFYING AND ADDRESSING GLOBAL TRENDS TO RESTRICT ACCESS TO SCIENTIFIC DATA FROM GOVERNMENT FUNDED RESEARCH

by Jerome Reichman, Duke University School of Law &
Paul Uhlir, The National Academies, U.S.A.*

The importance of public-domain data for scientific research is taken so much for granted—especially in the United States—that it is difficult to identify the precise boundaries of this domain, to describe its operations, and to evaluate the normative and legal infrastructure that supports it. Individual researchers know a little about the public domain as it affects their discipline, but very few understand how the overall system works.

In the United States, both science and innovation have been supported by a vast public domain for scientific data. In effect, scientific data has thus largely been a public good, which is available as essential inputs to both public and private researchers at the marginal cost of dissemination. In particular, the advent of digital telecommunications networks has made it possible to gather and share scientific data on a vast worldwide scale. Instead of physical repositories, data can be collected in networks of nodes and combined into virtual repositories open to world science. Internet links also have made it possible for scientists everywhere to access the same data collections, and efforts were being made to break down territorial barriers to sharing. Considerable progress has been made in collocating scientific data from all countries in both centralized and distributed repositories open to the world.

I. NEGATIVE TRENDS AFFECTING ACCESS TO GOVERNMENT-FUNDED RESEARCH DATA

The system that traditionally supported the public domain for scientific data depends

---

* This presentation is based on J. H. Reichman and Paul F. Uhlir, *A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment*, 66 LAW AND CONTEMPORARY PROBLEMS 315-462 (2003).

on a number of critical components or conditions that have recently come under attack. Let me review the basic components of the open access system in the U.S. as it was structured in the 1980s and early 1990s.

1) The U.S. government itself generated vast amounts of data, but it deliberately renounced any intellectual property rights in its own data; and these data were often made available in central repositories open to the world. (Other governments, however, do not necessarily follow the same policy).
2) The U.S. government also funded vast amounts of scientific research conducted at universities and other non-profit research institutes. Although the policies and practices of different funding organizations and types of research make sweeping generalizations difficult, many of the resulting data were eventually made available to other scientists under an open access policy supported by government funding agencies and by the sharing norms of science that prevailed in many disciplines.
3) Intellectual property law—as embodied in the classical patent and copyright paradigms—did not protect data as such, but actually consigned data to the public domain. Trade secret law protected information expressed as technical know how, even in disembodied form, but anyone was free to reverse engineer this know how, once embodied in commercial products, by honest means.
4) Publicly funded research results on the whole were not commercialized before the 1980s; universities did not generally seek intellectual property rights in their research results before the 1970s and federal agencies limited what universities could do with federally funded research results until 1980.

Unfortunately, in the past 25 years, all of these premises have buckled under a variety of economic, legal and technological pressures that have been placed on public-domain scientific data throughout the world. I will review some of the major negative developments in this trend and briefly discuss their implications for science.

1) The U. S. government has recently cut expenditures on data production by

transferring to the private sector a range of data production and distribution activities that were once publicly provided. Larger and larger portions of the public data 'commons' have thus been transformed into 'private monopolies' that do not adequately provide for public research on affordable and open terms.[1]

2) The government has also reduced the amount of money it spends to fund basic scientific research at universities, although the amounts it still spends are relatively high compared to other countries. Meanwhile, since 1980, the Bayh-Dole Act legitimated and further encouraged the growing tendency of universities to patent or otherwise protect and exploit their research results even when funded by government. More recently, these activities have expanded to securing both patents and copyrights in university-generated computer programs; to the patenting of molecular biology databases; and to the licensing of scientific databases as research tools on increasingly restrictive terms, that include restrictions on use and even grant back clauses and reach-through clauses claiming interests in future applications.[2]

Some leading universities in England have begun to protect databases so restrictively that, when scientists leave to take up positions at other universities, they are denied access to the very data they initially collected. These practices are reinforced by the European Union's strong exclusive intellectual property right in collections of data, which some universities in Europe are aggressively exploiting. This takes us to the next development, that is, changes in intellectual property law itself.

3) As I said at the outset, traditional intellectual property laws, especially copyright law, excluded facts and data as eligible subject matter. Even when facts and data were embodied in some copyrightable works of authorship, such as historical or scientific works, well-established limits and exceptions

---

[1] Reichman & Uhlir, 366-70; David, *Digital Boomerang*, at 12.
[2] Reichman & Uhlir, at 371; Rai & Eisenberg (2003).

to copyright protection allowed researchers to access, use and extract data for personal use and for public research purposes. However, recent developments in intellectual property laws in most developed countries now make it possible to assert and enforce proprietarial claims to virtually all the factual matter that previously entered the public domain once it had been disclosed.

A. The EU Directive on the Legal Protection of Databases (and Similar Sui Generis Regimes which have followed it).

The biggest change of all was the adoption, in Europe of a new exclusive property right in the very collections of data that traditional copyright law had left freely available from the public domain.

(1) This new exclusive right attaches to any database that is the product of substantial investment.

(2) The compiler obtains exclusive rights to extract or to reutilize all or a substantial part of the database.

(3) These exclusive rights, in effect, allow the compiler to control virtually all uses of the protected data, including follow-on applications of the same data in value-adding products.

(4) There are no mandatory public-interest exceptions to these exclusive rights in databases, as in copyright law, and there is a limited, optional exception for teaching. The Directive could thus prevent nonprofit scientists from incorporating an extract taken from a protected database into a new and different database.

(5) It is possible to take an insubstantial part of a protected database without permission, but it is risky to guess at what is an insubstantial part.

(6) Any new investment in updating or maintaining a protected database will extend protection of the whole database for continuing 15-year periods.

Hence, protection becomes potentially perpetual.

In a group of opinions handed down in 2005, the European Court of Justice seems to have cut back on database protection to a still unknown extent. In determining whether any given database qualifies for protection as a product of "substantial investment," the ECJ distinguished between investment for purposes of collecting the data (presumably not eligible in some or most circumstances) and expenditures for purposes of developing and maintaining the database as such (presumably eligible investment). However, once a database qualified for protection, the ECJ affirmed the high-protectionist effects outlined above.

How the Directive will actually affect science in any given country will thus depend on a number of imponderable variables. Nevertheless, we can say that this regime radically breaks with the historical limits of traditional intellectual property laws by protecting aggregates of information as such, potentially forever, without requiring any significant level of creative contribution. It affects all research by establishing a monopolistic barrier to the flow of upstream information, which has always been a free input into the information economy.

Moreover, these database protection laws could disproportionately affect scientists' access to and use of factual data historically in the public domain, primarily because they allow the scientist or the university to publish an article and still retain ownership of the data after publication. Similarly, this database right makes it possible for scientists or universities to apply for patents and disclose the supporting data in a patent application but still retain ownership of those data collections even after a patent expires.

B. Expansion of Copyright Law after the WIPO Copyright Treaty (WCT) of 1996

Even without an exclusive database right, copyright law has recently expanded in ways that can give very strong and virtually endless protection to databases in two ways. First, courts in the United States and elsewhere are increasingly inclined to see elements of creative selection or arrangement even in highly functional data sets and even for

value-adding purposes.

Second, the WIPO Copyright Treaty of 1996 (WCT) mandates anti-circumvention protection for all copyrightable databases transmitted through online digital networks. U.S. and E.U. laws implementing these provisions prohibit all unauthorized access to copyrighted databases transmitted online, even when the use for which access is sought would otherwise be permitted, e.g., for scientific research. The WCT does not require this result, but E.U. and U.S. law tend to override pre-existing exceptions in the online environment.

C. Other negative factors

Apart from intellectual property rights, digital technology itself enables database proprietors to restrict access to data tightly, even as it also makes it possible to share as never before. I refer to technological fencing and to one-sided electronic contracts, which make it possible to protect data delivered through online networks in absolute form, without any exceptions for science.

This result occurs when a technological fence forces the would-be user through an electronic gateway, where a one-sided electronic contract forces him to waive all user rights and privileges that copyright law or other laws might have given him. In effect, the contract becomes a privately legislated intellectual property right,[3] which recognizes no exceptions for science. When these technological fences and electronic contracts are supported by exclusive property rights—especially the database right—and by anti-circumvention measures, the compiler's power becomes virtually absolute and perpetual.

Meanwhile, as universities increasingly protect and exploit their research results, even when funded by government, the sharing ethos of science and the norms of open access to data have begun to break down in some fields. Because universities and scientists increasingly aim to commercialize their research products and profit from them, their willingness to exchange data, along with other research tools, has become seriously

---

[3] *See* J. H. Reichman & Jonathan Franklin, *Privately Legislated Intellectual Property Rights: Reconciling Freedom of Contract with Public Good Uses of Information*, 147 U. PA. L. REV. 875 (1999).

compromised.

For example, there is evidence that informal exchanges of data in biomedical research, between individual scientists and laboratories, usually at the prepublication stage, are breaking down with as much as 50 percent denials of requests for information.[4] At the same time, inter-university exchanges of data are subject increasingly to high transaction costs, delays, and a growing risk of anticommons effects,[5] that is, too many intellectual property rights and commercial interests make it difficult to put together complex data transactions. As relations between universities and industry become more intense, the universities tend to view each other as competitors, rather than partners in a common mission; and there is a growing ability of the industrial partner to impose the terms of exchange on the universities.

D. Conclusion

When all of these negative factors are added together—that is, the legal, economic and technological pressures on the scientific research commons—the threats to the conduct of science as we know it appear formidable. The new *sui generis* exclusive property right in collections of data poses a particularly serious problem. It will initially apply in some fifty countries affiliated with the E.U., and it is spreading to other countries through trade agreements.

Such a database right when combined with other economic and legal pressures could become the hub of a growing enclosure movement that could progressively fence off the public domain for scientific data. It could reduce the flow of data as an input into national systems of innovation, and into the international system.[6] Policy makers,

---

[4] Eric G. Campbell, et al. 2002. "Data Withholding in Academic Genetics: Evidence from a National Survey," 287 JAMA 473-80. Moreover, 10 percent of all post publication requests for additional information were reportedly denied.

[5] See, e.g., Stephen M. Maurer. 2001. "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, October 5-6.

[6] *See generally* Keith E. Maskus & Jerome H. Reichman, *The Globalization of Private Knowledge Goods and the Privatization of Global Public Goods*, 7 JIEL 279-320 (2004),

therefore, must take steps now to address these challenges and to ward off this threat of undue enclosure[7] in order to be able to exploit the new benefits of digitally linked databases.

II. A CONTRACTUALLY RECONSTRUCTED RESEARCH COMMONS FOR SCIENCE AND INNOVATION

The trends outlined above that favor commoditization of scientific data could elicit one of two types of responses. One response 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 confront the challenge by formulating a strategy that would enable the scientific community to take charge of its basic data supply and to manage the resulting research commons in ways that preserve its public-good functions without impeding socially beneficial commercial opportunities.

A. The Challenge for Science

Under the first alternative, the research community can join the enclosure movement and profit from it. Thus, both universities and independent laboratories 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 (that we already see in biotechnology today). 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 research community, as these same universities and research centers increasingly see each other as competitors rather

---

reprinted in INTERNATIONAL PUBLIC GOODS AND TRANSFER OF TECHNOLOGY UNDER A GLOBALIZED INTELLECTUAL PROPERTY REGIME (K. E. Maskus & J. H. Reichman eds. Cambridge Univ. Press, 2005).

[7] *See generally* James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 LAW & COMTEMP. PROBS. 33 (2003).

than partners in a common venture. Carried to an extreme, this competition of research entities against one another, conducted by their respective legal and technology transfer 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 basic 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. In approaching this option the community's assets are the formal structures that already surround government-funded data and the ability of the government 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 to react collectively to new information laws and related economic and technical pressures by negotiating contractual agreements between stakeholders to preserve and enhance the research commons.

In the United States, the government generates a vast public domain for its own data by a 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 made available at the marginal cost of dissemination, and it has been a major proponent and user of the Internet to disseminate its information as widely as possible. *In other words, it has deliberately made use of existing intellectual property rights, contracts, and information 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.*

Now that the research commons has come under attack in the United States and elsewhere, the challenge is not only to strengthen a demonstrably successful system at the

governmental level, but it is 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, should 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. 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 organizations to develop suitable contractual templates that can be used to regulate or influence the research commons.

Implementing these proposals will 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 briefly summarize our proposals for preserving and promoting the public-domain status of government-generated scientific data and of government-funded scientific data, respectively.

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 that exercise intellectual property rights in their own data collections.

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. A vast system of institutional mechanisms for the diffusion of scientific data as a public good—especially government-generated data—exists and continues to operate, and much government-funded data emerging from the academic communities continues to be disseminated through these well-established mechanisms.

In the European Union, a strong database right exists—as we have seen—and governments already exercise this right. Wise statesmanship would require these governments to renounce part of their rights to better promote scientific progress. Governments in all countries, moreover, should consider imposing contractual templates in their own relations with the private sector. When they license data to the private sector for exploitation, they should ensure that the private sector will not harm scientific activities when it exploits the data.

Contractual templates are thus needed to govern those kinds of relations and to help ensure the more efficient operations of the public-sector research community. These underlying contractual templates could implement the following research-friendly guidelines:

1. A general prohibition against legally or technically hindering access to the data in question for nonprofit scientific research and educational purposes;
2. A further prohibition against hindering or restricting the reuse of data lawfully obtained in the furtherance of nonprofit scientific research activities; 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.
4. We would also recommend the insertion of clauses dealing with misuse or abuse of any relevant intellectual property rights into all agreements between governments and third-party contractors.

In effect, we are suggesting that governments concerned about the preservation of a worldwide research commons should formulate their own database policies regulating their own data, with a view to promoting research without choking commerce. Governments from high-protectionist areas and from low-protectionist areas will also have to come to some sort of international understanding, perhaps a treaty, that will allow minimum protection of databases without requiring every country to adopt the highest

form of protection while keeping in mind the special requirements for the circulation of scientific data between countries with different levels of protection.[8]

C. Proposals for the Academic Sector

The primary focus of this presentation is on proposals for the academic sector, because so much of today's most scientifically important data are government funded and go through the universities. Therefore, they already benefit from a rudimentary regulatory regime. We suggest that science policy should treat data produced with government funds as a collective resource for research purposes and offer proposals that deal with how to do that.[9]

It is helpful to differentiate between two "zones" of government-funded data. The first is a zone of formally regulated data exchanges, for which the regulations are imposed by the funding agency and generally kick in at the time of publication. The second is a zone of informal data exchanges, which typically occur in the prepublication research phase, as well as in situations in which the terms of making data available are not formally specified in a research contract or grant. In Europe, the informal zone would also lie in the post-publication phase because of the existing database right. The ability of government funding agencies to influence data exchange practices will be much greater in the formal zone than in the informal one.

D. The Zone of Formal Data Exchanges (Academic Sector)

When no significant proprietary interests come into play, the optimal solution for government-generated data and for data produced by government-funded research is a formally structured archival data center also supported by government. Many such data centers have already been formed around large-facility research projects. 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

---

[8] See Professor Reichman's review of database protection in the international context, which proposes this type of treaty. J.H. Reichman. 2002. "Database Protection in a Global Economy," 2002 *Revue Internationale de Droit Economique*, pages 455-504.

[9] For detailed information on these proposals, see Reichman and Uhlir, op. cit., note 1, pages 425-456.

academics in different countries.

The traditional model entails a "bricks-and-mortar" centralized facility into which researchers deposit their data unconditionally. In addition to academics, contributors may include government and even private-sector scientists, but in most cases the true public-domain status of any data deposited is usually maintained. Examples in the United States include the National Center for Biotechnology Information, directly operated by the National Institutes of Health; and the National Center for Atmospheric Research, operated by a university consortium and funded primarily by the National Science Foundation. Hundreds of such data centers already exist.

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." Together, the network of nodes constitutes a virtual archive whose relatively small central office oversees agreed technical, operational, and legal standards to which all member nodes adhere. Examples of such decentralized networks, which operate on a public-domain basis in the United States, are the National Aeronautics and Space Administration's Distributed Active Archive Centers under the Earth Observing System Program and the Long Term Ecological Research Network funded by the National Science Foundation.

These virtual archives, known as "federated" data management systems, 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. "Conditional availability" means that the members of the network have agreed to make their data available for public science purposes on mutually acceptable terms, but the members also permit suppliers to restrict uses of their data for other purposes, typically with a view to preserving their commercial opportunities.

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 make these federated data management systems particularly responsive to the realities of current university research in areas of scientific investigation where commercial opportunities are especially prominent.

1. Basic Proposals

Several proposals are suggested for universities, and include the following:

(1) Develop interinstitutional agreements and cooperative institutional approaches to ensure unimpeded access to and liberal uses of scientific data and information in a not-for-profit research commons, while allowing for commercial exploitation of those resources in the private sector, when this is considered necessary and appropriate.

(2) Develop model contractual provisions for interuniversity and interresearcher relationships and for cooperative research with the private sector that protect access to and unrestricted use of publicly funded research data by not-for-profit scientists.

(3) Vigorously promote nonexclusive licensing by authors of their scientific articles to STM journals rather than assigning exclusive copyrights.

(4) Initiate and review pilot projects for certain disciplines or categories of data to test the results.

2. Ancillary Considerations

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. Particularly relevant in this regard are the open-source software movement that has collectively developed and managed the GNU/Linux Operating System and the Creative Commons, which seeks to encourage authors and artists to 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. Although neither of these models was developed with the needs of public science in mind,[10] 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.[11]

3. Further Analysis

We propose the contractual regulation of government-funded data in two specific situations:

(1) Where government-funded university generated data are licensed to the private sector, and

(2) Where such data are made available to other universities for research purposes.[12]

Case 1: Licensing Data to the Private Sector

This scenario presents the situation in which the possibility of "conditional deposits" of even government-funded data become most relevant. Now, the deeper question this raises is, how will universities and other research entities react to the new information

---

[10] Creative Commons has now undertaken to organize as scientific branch of operations. See _____.

[11] See pages 425 to 456 of Reichman and Uhlir, op. cit., note 1. We propose the contractual regulation of government-funded data in two specific situations: (1) where government-funded, university-generated data are licensed to the private sector, and (2) where such data are made available to other universities for research purposes.

[12] See pages 425-456 of Reichman & Uhlir, supra note 1.

laws, in which public research is going to have to operate, so as to make centralized data centers more viable and effective? After all, if you establish data centers, but the beneficiaries of government funding decline to deposit their research data because of their commercial interests, then we are not going to avoid erosion of the sharing ethos. So let us think a little bit about some more creative ways in which we might react to the legal and economic pressures that were previously mentioned.

My first point is that government should not just sit back and allow the spirit of the Bayh-Dole Act to apply to data collections in some uncritical fashion. The positive side of Bayh-Dole is that, in some sectors, particularly biomedicine, it has promoted the transfer of public research results to the private sector, and we want to keep the fruits of that success. The problem is that it has also generated unintended consequences and a whole slew of new questions, about secrecy, patent thickets, transaction costs, for example, that we need to think about. Do we want Bayh-Dole as it has been applied to patented inventions also to influence the commoditization of noncopyrightable databases? I think not. We want to arrive at solutions and protocols that enable government-funded databases to be commercialized without destroying their research value.

Data are not patented inventions. The policies and procedures applicable to one do not automatically apply to the other. Because data are always inputs to further upstream research efforts, we need especially to worry about the control of scientific data after publication that is almost certain to become a bigger problem the moment a database protection law is enacted.

Now, having said that, we must be careful not to undermine the spirit of Bayh-Dole. We do not want to discourage universities from privatizing their databases and extracting revenues from them when they can, in healthy competitive forms. But at the same time, it will not do to have the industrial partners of academic institutions dictating the terms on which government-funded data were made available for purposes of scientific research. That indeed would amount to a kind of reverse free riding.

On the contrary, what we find here is a real opportunity for government funders and

universities to develop their own database regime, a *sui generis* database regime that scientists like, rather than one that they hate, and apply that regime to the commercial areas in which government-funded databases are being exploited. Here, in other words, the scientific community has an opportunity to influence our industrial partners before they end up changing the public interest norms of academia.

Ideally, for example, the rules developed by universities and funding agencies could ensure that whenever government-funded data are used, any resulting follow-on applications built around those data could be freely made even in the private sector, provided that reasonable compensation were paid to the originating organization. This is what economists call a liability rule, instead of an exclusive property right (explain the difference). We have reason to believe that in the sensitive area of database protection, especially scientific databases, the public interest would benefit more from the application of liability rules than from an exclusive property right.

To take other examples, the universities and the funding agencies could impose sound legal standards regulating licensing terms, and prohibit abusive or unacceptable terms in licensing agreements when the database being licensed was government funded. In other words, we can develop agreed contractual templates that all universities would use when licensing government funded data to the private sector, with a view to facilitating general R&D uses of data. This would avoid a race to the bottom in which universities were out there trading away more and more rights of access to data for research purposes in order to attract better deals from the private sector.

Now, I will not go into further details here, but the point is that these mutually agreed contractual templates could themselves become a kind of model database law that does balance public and private interests in ways that the legislated database protection laws may fail to do. Unless science itself takes steps of this kind, there is a risk that, under some future extension of Bayh-Dole, the private sector will impose its own database rules on government-funded data products developed jointly with the universities.

Case 2: Inter-University Licensing of Data

Now, let us take one more big step forward. Whatever wisdom or lack of wisdom there may be in our idea about influencing the private sector through contractual templates to be imposed on transactions affecting government-funded data, it ought to be clear that the one thing science policy absolutely must do is to establish similar contractual templates to regulate inter-university relations with respect to government funded data. That is a responsibility we cannot shirk. In inter-university relations, the public stake is very great, not only because the data has been generated for scientific purposes by public expenditures in the first place, but also because the progress of science will itself depend on continued access to--and further applications of--the data that have been generated.

That invokes a subtext of Bayh-Dole itself that merits careful consideration. Within Bayh-Dole there are clauses that allow for march-in rights, for compulsory licensing, and for the preservation of the public interest. However, these march-in rights have never been exercised, or at least successfully exercised.[13] (Cook-Deegan mentions one unsuccessful attempt to exercise them in the *CellPro* cases; another case concerning AIDs drugs that was pending before the NIH was also dismissed). We need to build on the public- interest aspects already recognized in Bayh-Dole and make them more functional, by working together, by getting the universities, the funding agencies, and the scientific bodies to develop their own set of rules regulating government-funded data sets.[14]

Consider that if a strong database protection law were enacted in the United States, and if Bayh-Dole were literally applied to the new intellectual property right, it would just simply pass those strong exclusive rights to data through the system, to universities and academics, who would then own all this government-funded data. Without any agreed restraints on how they treat that data, university technology transfer offices could treat databases just like patented inventions, even though we know that the potential

---

[13] See Rai & Eisenberg, above n. ___.
[14] For proposals to reform the Bayh-Dole Act directly *see id.*

impact on research would be enormously greater. One can immediately predict a race to the bottom if universities tried to maximize revenue from these rights. We can even envision a bizarre scenario in which the universities exercised their right to contractually override any modest research exemptions that happened to appear in some officially legislated bills!

Our goal instead is to persuade universities and governments agencies to develop their own database regime to cover scientific uses, whatever else they want to do with their data in the commercial sector. The first principle, if you accept our invitation to negotiate our own contractual templates, is that nothing the universities do should derogate from the customary and traditional uses of scientific data by the scientific community. Then, having accepted that principle, the negotiators should go farther and delineate a positive, agreed set of contractual templates that regulate inter-university uses of government funded data on fair and reasonable terms and conditions. Absent such a proactive approach, we fear a slow unraveling of the traditional sharing norms and a race to the bottom.

What kind of norms would emerge from an inter-university pact supported by the funding agencies? And what are the difficulties of negotiating such a multilateral pact of this kind?

Consider that we are now talking about scientific uses of government-funded data in an academic context. So we would all immediately agree that the open access and sharing norms of science ought to be a kind of foundational institution. One wishes it were that simple, however. In reality, universities have already demonstrated a tendency to over-restrict patented research tools, including databases, even for other academic purposes, under Bayh-Dole. There is reason to believe that this restrictive approach would be applied to commercially valuable databases generally, especially if the university technology transfer offices were invested with a *sui generis* intellectual property right in collections of data.

So the universities are going to be very conflicted about what to do with their databases. Consider that any effort to lock the universities into a regime that guaranteed

access to and sharing of government-funded data would require these same universities to convert their statutory exclusive property rights into a kind of voluntary rights pool for scientific purposes. Conceptually, this seems like a rather elegant solution. From the intellectual property literature, we are familiar with the advantages and synergies that may come from pooling. But from the antitrust literature, we are also very much aware of the dangers that reside in pooling, the possibilities for manipulation, discrimination, defection and holdouts, and the centripetal forces that can rip pools apart when one party misbehaves and others emulate that strategic behavior.

Of course, our preference would be that inter-university exchanges of government-funded data should operate on the basis of a true research commons. The conceptual framework prodding universities to do that would be for them to conclude that, in the long run, they would gain more from free access to each other's databases than they would lose from the particular gains resulting from commercial exploitation of any particular research tools that they happened to own at any given moment. That is the core concept. I believe it is an accurate concept, but we have no empirical verification, so it would be an act of faith.

Certainly, if we can convince the research universities to embrace that concept, then it is worthwhile to seek a negotiated solution that would preserve the open access norm across the board. But I can immediately tell you what will happen if we are not careful, I can predict the first problem that will arise even if all the universities got up today and said, yes, Professor Reichman, we buy into this concept, we will adhere to a true public-domain principle. Tomorrow morning, first one and then another technology transfer official would appear at our door saying, however, Professor Reichman, this particular database has been funded with some private data in addition to government-funding; it is the product of mixed funding sources. It does not just contain public data, it contains some private data, too, so naturally we are entitled to an exception from the general rule. We would accordingly need to deposit these data on a more restricted basis; we need to make a *conditional deposit* so as to preserve the rights of our private-sector partners.

They would have a valid point. Would a drive for totally unconditional deposits of

government-funded data succeed, or would too many universities hold out for a two-tiered system, in which conditional deposits were also allowed? Can we succeed in maximizing access to government-funded data without a two-tiered system, in which some data could be conditionally deposited? We fear that once a database protection law was enacted, and Bayh-Dole were factored in, along with the practices of the technology transfer offices as they now stand, it might become very difficult to persuade the universities both to agree to, and actually enforce, a true public-domain model for all government funded databases.

In that case, we could experience a lot of unraveling of the sharing ethos. We could quickly precipitate from a best-case scenario in theory to a worst-case scenario in practice, as universities started bargaining to impasse and we encountered more and more of all the problems—especially anticommons and hold-out problems—that we already know about with regard to present-day biotech databases. So we believe that the worst-case scenario is so bad, in which everyone starts bargaining in every possible way, and the commercializing pressures would be so great in the presence of a database intellectual property right, that steps must be taken to head off the collapse of the best-case scenario by judiciously allowing conditional deposits of government-funded data in specified, agreed situations.

Remember, we are not talking about a few hundred patents now. We are talking about every single government funded university data product that has potential commercial value as a research tool and in other ways, too. We fear that, if the scientific establishment cannot achieve perfect harmony on a single-tiered model, then proper data management policies may require a second-best solution: namely, a negotiated contractual template allowing universities to make conditional deposits and even to seek compensation from other universities for certain data tools on certain agreed standard terms and no other terms.

Now, immediately you will observe that this proposal reveals the horns of a dilemma. One side of the dilemma is this. If no two-tiered model exists, then everything depends on the success or failure of the best-case, true public-domain model. If that fails,

however, you go directly to the worst-case scenario. But the other side of the dilemma is this. If you negotiate an acceptable two-tiered model for a contractually reconstructed research commons, one that we could all live with, then, of course, there are going to be pressures on universities to put more and more of their data into the conditional deposit regime and thereby avoid the true public-domain preserve.

Can the funding agencies press the universities to limit recourse to conditional licensing? We don't know the answer to that, but we doubt it. Notice that a conditional licensing model responds to market forces, and it has one positive feature that is worth emphasizing. It solves that problem of the borderline where government–funded data are mingled with privately contributed data. That is, so long as there is a second agreed set of contractual templates allowing conditional deposits, then a given university can no longer say, I want a still better deal because I have a private partner. The private partner will then be held to the second set of conditional templates and nothing else, because that is the agreed pact. The private partner cannot override agreed terms that provide access to data for public research purposes.

We don't know which is the best solution. The object is to get the most data revealed, and to avoid both secrecy and transaction costs as much as possible. Possibly different scientific communities would prefer different solutions. Astronomy would certainly opt for the true public-domain model and nothing else. On the other hand, certain communities are already so close to the worst-case scenario that they would immediately benefit from an agreed contractual template that permitted conditional deposits in order to encourage more disclosure. We don't know which system will promote the most disclosure, these are things that we have to find out in practice, and we invite you to help us think about them.

THE INFORMAL SECTOR

I will not have time to discuss the additional proposals that we have in mind for the zone of informal data exchanges between single laboratories or investigators. In some ways, this zone has become even more critical to scientific progress than the formal structure. But I will just tell you in two sentences what the project is that we envision.

What do we want to do? We want to realize digital opportunities to link distributed databases into functioning virtual archives. In a word, as Harlan Onsrud will talk more about and has done a lot of pioneering work on, we want to "napsterize" the whole disaggregated bunch of laboratories out there (that Stephen Hillgartner so ably described). One way to do that is to borrow some lessons from the open-source software movement and from the Creative Commons Initiative, i.e., to negotiate our own database arrangements, agreed contractual templates for exchanging data that everybody can live with, and then to put those together and create a kind of voluntary research commons with different options available. This solution would require a two-tiered option, but one that could not make matters worse than they are at present. It could instead avoid destroying the fragility of the informal system of data exchanges—the phrase used by Paul David which haunts me—because it is a very fragile system, indeed, and we don't want to destroy it.

Now, even though I cannot say more about this, many other speakers at this conference are going to tell you about all the interesting projects that are going on. You will also hear a talk about the Creative Commons organization, which is actually beginning to do something very much like this in the scientific domain, and we can learn a lot from them.

Finally, let me remind you, though we have no time to speak about or even focus on it at this conference, we really have to worry about making private sector data more available for public research. We think that this contractual template approach could do it. We urge you to think about that later on, too.

Thank you for your attention, and I'll take any questions.