

New Models of Information Production and Management in Public Research: Legal, Economic, and Science Policy Considerations

Paul F. Uhler
puhler@nas.edu

Introduction

The boundaries between the open and closed segments of public-sector information regimes are dynamic and shift over time and place. Some of the factors that lead to significant changes are global or regional, while others are specific to a nation's political system or culture. Among the global and large-scale regional factors that have promoted greater openness in public-sector information over the past decade are the collapse of the Soviet empire, the rapid adoption of digitally networked information and communication technologies (ICTs), and the continuing "globalization" of the world economy.

The disintegration of the Soviet Union and its geopolitical sphere of influence has resulted in great wave of liberalization in public information regimes. The wholesale abandonment of communist dogma and the related tight state control over all aspects of information activities has given rise instead to new constitutional and statutory guarantees of freedom of expression, freedom of information, and the proliferation of competing media and other sources of information production and dissemination. Although most of the newly independent states of the former Soviet Union have not yet enjoyed these freedoms, Russia itself, Eastern Europe, and many developing countries have enacted freedom of information laws and constitutional amendments favoring citizen's rights of access to government information and other information rights previously denied.¹

The explosion of access to the Internet and ICTs over the past decade, especially in OECD countries, has enabled governments to broadly and cheaply make available public information products and services. This trend has resulted in numerous "e-gov" initiatives at local and national levels throughout the world. Growing access by citizens to digital

¹ For a collection of these statutes worldwide, see <www.freedominfo.org>.

networks also has placed pressures on governments to become more open with their information, particularly in those countries with previously (or currently) closed information regimes.

Certainly the globalization of the world economy has both spurred a normative standardization of government information practices and been greatly facilitated by the growth of the Internet on many levels. These overarching political, technological, and economic trends have both required and made possible greater transparency in governance and stimulated more open disclosure and availability of many types of public information.

At the same time, there have been strong forces driving governments to further restrict the information they produce at public expense. Many of these forces emanate from the contrary or reactive interests inherent in the very same factors and trends that have promoted openness in public information. For example, the national security based restrictions on potentially sensitive public-sector information that were relaxed at the end of the Cold War are being resuscitated as a result of the increasing threats from global terrorism and from the proliferation of weapons of mass destruction. The heightened security concerns have led to the enactment of new legislation and to the revision of government regulations and policies further tightening access to public information considered sensitive or potentially harmful.²

The emerging information economy that has grown vigorously as a direct consequence of the revolution in digitally networked ICTs also has led to pressures to privatize public-sector information activities and to commercialize many government generated and government funded information resources. Although the shift of economic activities from the public to the private sector has been underway over several decades, it accelerated significantly after the late 1980s, following the collapse of the communist bloc. This

² For example, see Adam Clymer, "Government Openness at Issue as Bush Holds on to Records," *The New York Times*, January 3, 2003, p.1, 16, for a summary of such restrictions in the United States following the September 11, 2001 attacks, including a chart of the number of documents officially classified every year from 1991 to 2001.

larger trend toward privatization of government functions also extended to information products and services in most countries. Moreover, as digital information has emerged as an increasingly valuable resource during this same time period, government entities have sought to recover their costs of producing and managing it, in many cases selling it back to their citizens at commercial rates. Both the privatization and commercialization of public information products and services has diminished the access to and reuse of such information. A similar trend has occurred with regard to public research and the information produced from that.³

Finally, the capability of the Internet to make available information products to millions of potential users openly, instantaneously, and at zero marginal cost has resulted in reactive legal responses by established firms in the information industry that see such capabilities as a direct economic threat. These large economic interests have lobbied successfully at both the national and international levels for the adoption of protectionist intellectual property (IP) treaties and statutes.⁴ They also have brought into widespread use highly restrictive information licensing practices—enforced by increasingly effective digital rights management technologies—that have skewed the social and economic balance of rights in information. New IP laws and private contracts have broadened, lengthened, and deepened the scope of legal protection in information goods generally, with still poorly understood effects—both intended and unintended—on the information economy.

This legal protectionism in the information sector has been fortified through the globalization process under the World Trade Organization and the 1994 Agreement on Trade Related Intellectual Property Rights (TRIPS), as the economically developed economies have sought to protect their industries and promote trade on favorable terms. Not only has this tightening of the IP regime at both the global and national levels led to reductions in the rights of user access to and use of private-sector information sources,

³ Cite, e.g., US and UK statistics

⁴ For texts and background summaries of new and proposed treaties and statutes in copyright and neighboring rights, as well as in other IP law domains, see the World Intellectual Property Organization Web site at <www.wipo.int>.

but has produced a similar effect in the public information sector. In those countries that apply IP protection to public information, such added restrictions of course directly reduce the public's rights of access and use. However, even in those jurisdictions or information domains in which IP protection of public information resources is not allowed, the increased legal protection has indirectly reduced the unrestricted availability of government information by increasing incentives and pressure from the private sector to gain control over those public resources for profit.⁵

Consistent with—and in spite of—these larger trends and forces that continually shift the boundaries of the open and closed segments of public information regimes, there are many legal, economic, and other public-policy reasons that support the placing of most information produced by governments as public goods in the public domain⁶ under conditions of open availability and unrestricted reuse. First, a government entity requires no incentives from intellectual property laws to create information, unlike individual authors or investors and publishers in the private sector. Second, the information has been produced at the expense of the taxpayer. The moral rights in that information may be said to reside with the citizens who paid for it, rather than with the state entity that produced it on their behalf. Third, democratic values and transparency of government are undermined by restricting citizens from access to and use of public information. Citizens' rights of freedom of expression also are compromised by restrictions on redissemination of public information, and particularly of factual data. It is no coincidence that the most repressive political systems have the least amount of available information and the greatest restrictions on expression. Finally, there are numerous positive externalities—particularly through network effects—that may be realized through the open dissemination of public-

⁵ For a discussion of the increasingly protectionist information law regimes and their potential effect on public-sector scientific data activities, in particular, see J. H. Reichman and Paul F. Uhler, "A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment," Part III, "Pressures on the Public Domain and Their Implications," *Law and Contemporary Problems*, Duke University School of Law (forthcoming 2003).

⁶ Public-domain information may be characterized as "sources and types of data and information whose uses are not restricted by intellectual property or other statutory regimes and that are accordingly available to the public for use without authorization or restriction." Paul Uhler, "Session I Discussion Framework: The Role, Value and Limits of S&T Data and Information in the Public Domain," in *Proceedings of the Symposium on the Role of Scientific and Technical Data and Information in the Public Domain*, Paul F. Uhler and Julie M. Esanu, eds., National Academy Press, Washington, DC (forthcoming 2003).

domain (as well as proprietary) data and information on the Internet. In many cases, such benefits are not readily quantifiable and extend well beyond the economic to social, educational, cultural, and good governance values.⁷

All OECD countries designate at least some categories of government information, such as judicial decisions and legislative acts, in the public domain, although the scope of unrestricted information varies considerably among them.⁸ With the notable exception of the United States, however, other nations apply copyright (e.g., Crown copyright in the U.K.) and related forms of IP to government information that is not expressly designated as belonging in the public domain, although they may nonetheless make some of their proprietary information openly and freely available. In addition, all countries have other well-established categories of information that are kept secret or restricted under specific policies and statutes, based on the aforementioned national security concerns, and on the protection of personal privacy, confidentiality, and private IP rights.

It is useful to compare and contrast some of the practical effects of open and proprietary public information regimes. For example, all OECD countries have public-sector meteorological offices and activities. Both the U.S. and the E.U. (as well as Japan and Russia) also operate public-sector meteorological satellites. All of the governments choose to maintain their metsat systems under government control, even though both the technology and the commercial market for the resulting information are sufficiently developed to support a private-sector operator. They do this to maintain public control over a technology and information system that has very significant public safety, economic, and national security implications. However, the U.S. and the E.U. apply very different information regimes to the resulting data and information from their metsats and other meteorological data systems.

⁷ Ibid.

⁸ For a somewhat dated description of laws regarding access to public information in E.U., see the European Commission, “Public Sector Information: A Key Resource for Europe, Green Paper on Public Sector Information in the Information Society,” COM (1998) 585. For Freedom of Information laws that currently are in force in all countries, see www.freedominfo.org, *op cit*, note 1.

In the United States, the government makes the data available freely and openly by direct transmission from its satellites to anyone who has a receiver anywhere in the world, distributes much of its processed weather information freely online, and makes available its archived data at marginal cost of dissemination in the public domain from a public data center. This policy has a multiplier effect in which the nation (and the world) derives greatly increased returns on the public investment. By making the resulting data freely available in real time the government provides information not only in support of its narrower public safety mission, but produces the raw material for a value-adding weather information sector, for further research and applications, for education, and for countless other beneficial downstream uses by sophisticated businesses and ordinary citizens alike. Indeed, the private-sector weather information industry in the U.S. now generates over \$500 million in business each year, developing innovative new information products, employing thousands of people, and generating significant tax revenues for the public treasury. And, when the government places its data freely on the Internet, the resulting network effects can make these benefits exponential.⁹

Contrast this approach, if you will, with the one taken by the governments of the European Union and their meteorological agencies, which seek to recoup as many of their data management operating costs as possible at the original point of dissemination. Instead of making available such information as freely and openly as possible, they charge commercial rates, enforced with strong proprietary legal restrictions. Instead of supporting the growth of a value-adding industry, which is all but non-existent in Europe, the data provide much more limited economic returns. Instead of disseminating the data for any researcher to use as desired, only some of the data may be available for unrestricted use. Instead of providing myriad unfettered opportunities for social, educational, and serendipitous applications, the data are dribbled out at high costs in small amounts, intended for end users only and only to those who can pay.

⁹ See, for example, Peter Weiss, “Potential Effects of a Diminishing Public Domain in Environmental Information,” in Proceedings of the Symposium on the Role of Scientific and Technical Data and Information in the Public Domain, *op cit*, note 6, providing economic statistics regarding the benefits of

This false economizing generates some modest economic returns for the state by offsetting a part of the operating costs of the public activity. But this is at the expense of countless lost socio-economic opportunity costs. The result is the worst of both worlds: the inherent inefficiency of a subsidized public monopoly operating outside the free market and stifling private-sector development, while behaving as an ersatz commercial entity, incurring all the deadweight social costs that might otherwise be avoided if the data were openly available in the public domain in the digitally networked environment.¹⁰

Although a comprehensive comparative analysis of open and restricted public information regimes is beyond the scope of this article, the foregoing example illustrates some of the relative merits of an open, public-domain approach and one that places proprietary restrictions on access to and use of public information.

Public-Sector Scientific Information

Publicly funded research itself produces a vast array of information that constitutes a major category in the larger public-sector information regime. Public-sector scientific information is directly influenced by, but also in many ways separated from, both the larger global trends and forces and the general public information regime. There are two key factors that separate public research activities from other public (and private) information sectors. The first is the highly specialized and esoteric nature of much research and the data and information it produces. This imbues the conduct of research with a substantial degree of distance and autonomy from more prosaic economic pursuits. It also allows research institutions and researchers themselves a greater measure of freedom to self-regulate their activities, not least in how they handle their information resources.

The second factor, which arises in part from the first and in part from the intrinsic nature of the research enterprise itself, is the existence of a strong set of norms that structure

this open data dissemination policy for meteorological satellite data in comparison with the policy of commercialization and proprietary protection of such data by government agencies in the European Union.

relationships within the broader context of public science, as well as within more tightly integrated discipline-oriented sub-communities. The general norms and values of public science in government and academia, particularly in areas that may be characterized primarily as basic research, differ considerably from those of information producers in the private sector and even from those in other segments of the public sector. Public research is largely a communitarian, cooperative system.¹¹ It is founded on freedom of inquiry, sharing of data, and full disclosure of results by scientists whose motivations are typically rooted in intellectual curiosity, the desire to influence the thinking of others about the natural world, peer recognition for their achievements, and promotion of the public interest.¹²

Public science, while connected to the market economy, operates largely outside it. Public-sector research and the data and information it produces in many cases are not driven by practical objectives or direct applications that otherwise are required in private-sector research. Rather, it is the findings from such upstream public research activities that make possible the downstream innovations, which are properly incentivized and converted into practical commercial applications with intellectual property laws. Indeed, the very fact that certain research is publicly funded implies that it is an area of basic inquiry in which commercial private-sector firms have little incentive to invest, because the profits are too uncertain, too far in the future, or too difficult to capture. For society to derive the maximum social and economic benefits from public research requires the broadest possible access to the results, at the lowest possible cost, with the fewest inhibitions on their reuse.

As Michael Polanyi pointed out some 40 years ago, the community of scientists—which he called the “republic of science”—is a body politic that is organized around a set of

¹⁰ Ibid. For a more general comparison of the effects of these two policy approaches, see PIRA International Ltd., “Commercial exploitation of Europe’s public sector information,” European Communities, 20 September 2000.

¹¹ See, e.g., Robert K. Merton, “The Normative Structure of Science,” in *THE SOCIOLOGY OF SCIENCE*, Robert K. Merton, The University of Chicago Press (1973), pp. 267-278.

¹² R. Stephen Berry, “Is Electronic Publishing Being Used in the Best Interests of Science? The Scientist’s View,” *2 Int. J. Mol. Sci.* 133-139 (2001).

principles similar to the market economy, but outside it.¹³ Like the larger economy, the republic of science is composed of a large number of independent actors and initiatives that are nonetheless coordinated and interrelated. Just as the price structure provides the “invisible hand” that guides the operation of the market economy, the disclosure of research results provides the invisible hand in the research enterprise. The currency of the researcher consists of the ideas that perceptibly influence the thinking of others. But unlike the exchange of private goods in the market economy, the outputs of public research are public goods, which are not exhausted through consumption but actually *increase* in value the more that they are used. And. Although competition at the individual level motivates scientists to excel and publish, the republic of science is guided by the communitarian ethos. Science can only succeed by building on the results of others in the community, which must be openly disclosed and communicated.

These norms and values of the republic of science thus intersect with and closely parallel the legal, economic, and public policy reasons for placing the broader public-sector information in the public domain as public goods. Despite this confluence of values and goals, however, it is important here to make one key distinction concerning the information produced by scientists, whether in government agencies or academia. While most do not conduct their research for profit or to make money on the publication of their results, they do seek recognition and attribution for their contributions to their field and to the global knowledge base. As such, they require some assurance of authenticity and integrity of their work, typically provided through copyright. Unlike the publishers who disseminate their work, however, they do not seek to enforce the copyright for economic gain or as a barrier to access and use by others.

Peer Production and the Open Access of Information in the Republic of Science

Although this normative and value structure of public science predated the revolution in digitally networked ICTs, it makes it ideally suited to experiment with and exploit the capabilities of those technologies, which themselves facilitate open, distributed, and

¹³ Polanyi, “The Republic of Science,” 1 *Minerva* 59 (1962). The term “republic” itself is derived from the Latin words, *res publicum*, or “public thing.”

cooperative uses of information. It is no coincidence, therefore, that the emergence and early institutionalization of many new models of virtual knowledge-based communities and related means of information production on an open basis have occurred in public-sector science. Examples may be found in all phases of the information chain, from production to management to dissemination and archiving. In information production, they include collaboratories for virtual experiments, the construction of virtual observatories, the production of open-source software, open journals, and hybrid Web sites. In the management of information there are automated citation indices and open peer review processes. And for dissemination and archiving there are federated data management networks and open archives. Taken together, these emerging capabilities represent a broader trend on the Internet toward both formal and informal peer production of information in a highly distributed, volunteer, and open environment. Such activities are based on principles that may be more accurately characterized as intellectual commons, rather than intellectual property, and that reflect the communitarian ethos of public science. Table 1 provides a partial comparative list of characteristics of print and digitally networked information, which helps to illuminate some of the marked differences between these two paradigms and to explain why distributed, volunteer, and open information models are being successfully adopted within the republic of science.

The traditional capitalist model for the production and distribution of goods, including information products, is based on the principles of market economics and on hierarchical corporate institutional mechanisms. The success of both of these forms of information production and distribution requires well-defined and enforceable property rights to control the resources and outputs.¹⁴

In the information sector, the new characteristics of digitally networked information have changed some of the key premises on which the property rights under the print paradigm were based, and challenged the control and enforceability of those rights. In particular, the simple and low-cost ability to copy and re-disseminate proprietary information

¹⁴ Yochai Benkler, "Coase's Penguin, or, Linux and the Nature of the Firm," 112 Yale L. J. 369 (winter 2002-2003).

TABLE 1: Key Characteristics of Print and Digitally Networked Information

PRINT	DIGITAL NETWORKS
(pre) Industrial Age	post-industrial Information Age
fixed	transformative
static	interactive
rigid	flexible, extensible
physical	“virtual”
limited content and types	unlimited contents and multimedia
local	global
linear	non-linear, with time/space collapsed
high marginal distribution costs	zero marginal distribution costs
distribution physically difficult, slow	effortless, immediate universal distribution
copying cumbersome, not identical	copying easy and identical
single user	multiple, concurrent users
non-interactive	interactive

instantaneously and globally has made it easy for unscrupulous users to misappropriate an author’s or publisher’s proprietary information in market-destructive ways. This has led information rights holders, especially publishers, to seek to reassert control and enforceability of their proprietary rights through the various restrictive legal and technological mechanisms mentioned above.

At the same time, however, we have seen the emergence of a parallel trend of information production and dissemination in both the public and the private sectors based

on business and organizational models of reduced proprietary controls over information. The most obvious and pervasive manifestation of this is the emergence of millions of open Web sites that contain enormous amounts of all kinds of copyrighted, proprietary information. Many companies now give away information that they might have withheld or sold prior to the implementation of the World Wide Web. They do this for business reasons such as a loss-leader for other directly or indirectly related products and services, as an attraction for customers to Web sites that are paid for by advertising revenues rather than by the direct sale of content, or purely for publicity.¹⁵ Although most of these types of open information production and dissemination activities do not replace traditional publishing models, they have resulted in a broad perception by the general public that information on the Web is free and unprotected, much to the well-known dismay of music, film, and other commercial publishers.

Most interesting of all in the context of this article, however, are new forms of information production, management, and dissemination in both the private and public sectors that are based on non-market, non-corporate, and non-proprietary principles. Referred to generally as “peer production” of information, these approaches make use of the distributed interconnections on the Web and alternate incentive motives that are not dependent upon direct economic returns or intellectual property protection for success. Instead, they are based on non-economic incentives or indirect economic returns, and on volunteer, community-based actions that operate as intellectual commons, rather than as closely guarded property. There are a number of other key aspects of peer production of information that can make it a more efficient and preferable mode of organization to the traditional market-based, hierarchical model.¹⁶

The first is the need for a sufficiently compelling common objective that serves as an incentive to attract enough volunteer labor to make it work. This, of course, will vary

¹⁵ For a description of new business models and trends in the information sector that take advantage of the different characteristics of digital networks, see Varian and INFORMATION RULES (1999) [need to complete cite].

¹⁶ Based on elements identified by Benkler, op cit, note 14, and Charles M. Schweik and J. Morgan Grove, in “Fostering Open-Source Research Via a World Wide Web System,” available at <www.pamij.com/5_4/5_4_2_opensource.html>.

depending on the nature of the problem to be addressed and other factors discussed below. As already noted, the incentives in the republic science are not based on monetary rewards, as much as by the desire to influence the thinking of others, to advance knowledge, and to serve the public interest. Thus for scientists, the incentive to work on a volunteer, distributed information project will often depend on the intellectual stimulus it poses and the relevance of the topic to ongoing work. However, there are notable examples of peer production of information outside research as well.

The phenomenal success of the Linux operating system or the Apache Web server software as part of the free or open source software movement is particularly noteworthy. Here, thousands of computer scientists and software engineers have donated their time and labor to help develop and improve major software applications that have had a significant economic impact and value. The motivations for participation in the open software system have been largely attributed to the “hacker culture”—the interest to develop new and better tools in competition with Microsoft, where the market-based economic conditions for successful product competition did not exist. The ability to receive some attribution and recognition for contributing incremental advances in this effort within the hacker community has also been important, however.

Another relevant factor in the incentive structure for peer production is associated with the large scale of activity. If a project can attract the participation of thousands of volunteer rather than a few dozen or hundred, the personal cost of participation by each volunteer becomes quite low and the incentives to stimulate contributions are greatly reduced. An example of this is the distributed proofreading function of Project Gutenberg.¹⁷ This project converts old paper texts that are in the public domain into digital formats to make them globally available online. A related distributed proofreading site allows volunteers to proof the converted text in small increments, so that this painstaking and mundane task can be accomplished quickly and easily by people who may have a special interest in a certain book.

¹⁷ Example provided by Benkler, *op cit*, note 14.

Another example in the research area, but not relying on the participation of researchers themselves, is the NASA Mars Clickworkers project. This was an experiment to see if public volunteers, each working only briefly or occasionally, could perform routine scientific analysis that would normally take a scientist months to accomplish.¹⁸

Volunteers were asked to mark craters on maps of Mars, classify craters, or search the martian landscape for honeycomb terrain. In the first half year of operation over 85,000 volunteers visited the site, making almost two million entries, including redundant entries of the same craters, which were used to average out errors. An analysis of the quality of the markings showed that the automatically computed consensus of the large number of “clickworkers” is virtually indistinguishable from the inputs of a Ph.D. geologist with years of experience in identifying Mars craters. Although the question of adequate incentives in the peer production of information needs to be studied further, these and other examples illustrate that the threshold for active participation in a purely volunteer setting on the Internet is not high and in fact may not even require formal recognition or attribution to achieve certain desired outcomes.

The next three elements of peer production of information—trust, shared norms, and established rules of engagement—may be called the building of social capital.¹⁹ These are attributes that participants typically need to bring to their group interactions in recurring activities.

Trust can initially be developed by assembling a core group of collaborators who either may know each other or at least know something about each other’s work and institutional affiliations. As a project develops over time, the results and reputation of the project will presumably serve to build trust among the expanded group of participants. The open availability of the results is critical, however, if the objective is to involve many participants beyond the initial core group of collaborators.

¹⁸ Ibid.

¹⁹ Schweik and Grove, *op cit*, note 16.

Shared norms that support the open, distributed, and volunteer activity are of course essential as well. A key norm is one of reciprocity and cooperation. Another is generally accepted standards. Both of these are especially important if the peer production group is relatively small. Certainly here the norms and values typically associated with public science, as discussed above, are supportive of such endeavors. However, in larger-scale projects such as the NASA Mars Clickworkers or Project Gutenberg proofreading, these factors become somewhat less important.

Finally, an established system of rules under which the cooperation is to take place is also essential to the successful development of social capital. The strong default rules of individual ownership that are provided by the IP law regime need to be expressly mitigated in order to avoid conflicts of rights and subsequent legal liability claims. This topic itself could be the subject of an entire article, however. The key concept here is the use of licenses that derogate from the full IP rights that otherwise automatically attach to information projects.²⁰ Such public use licenses establish a type of conditional public domain or copyleft regime of reduced property rights, but that still are sufficient to prohibit the subsequent exclusive capture or misappropriation of the information produced. The idea is to promote open access and use in a socially productive manner.

Modularity and granularity refers to the attributes that are necessary for distributed, volunteer peer production of information to be a viable alternative.²¹ This means that the information product or activity must be composed of sufficiently small and discrete parts or functions to make autonomous, individual contributions possible. Thus an individual article cannot be coherently written by a group of strangers working autonomously, but a virtual archive of individually authored articles in some discipline or thematic area can be. Similarly, a database of autonomous observations collected independently according to some commonly agreed protocols can be assembled.

²⁰ See, for example, the licensing templates proposed by the Creative Commons at: <http://www.creativecommons.org>.

²¹ Benkler, *op cit*, note 14.

Finally, the costs of managing the project and integrating the disparate contributions must be reasonable and not excessive. This, of course, will depend on the nature of the product or activity, its intended use, and other factors specific to the project, including the modularity and granularity of the contributions and the level of expertise and quality necessary.

The communitarian ethos of the republic of science; the social, economic and political principles that favor open availability of government information in the public domain; and the interactive and integrating capabilities of global digital networks are together combining to enable exciting new opportunities for collaborative production, management, and open dissemination of scientific, and other, types of information. These opportunities are leading to a “re-intermediation” of scientific information outside the institutional and market-based mechanisms that arose under the *ancien regime* of the print paradigm. Although many examples of such activities have been sprouting up all over the Internet in recent years, the underlying functions and dynamics of these different approaches remain poorly understood. Certainly, this article, and the few that have been written beyond the open source software movement, have only begun to analyze this phenomenon, which will merit much greater attention in the coming years.