

DATA SHARING, LATENCY VARIABLES, AND SCIENCE COMMONS

Jorge L. Contreras[†]

TABLE OF CONTENTS

I.	INTRODUCTION	1602
II.	COMMONS AND LATENCY	1609
A.	COMMONS THEORY BACKGROUND	1610
1.	<i>Physical Resource Commons</i>	1610
2.	<i>Commons of Information</i>	1611
B.	THE TEMPORAL DIMENSION OF INFORMATION COMMONS.....	1613
1.	<i>Dynamic and Static Commons</i>	1613
2.	<i>Knowledge, Rights, and Timing in Information Commons</i>	1614
a)	The Dynamic Character of Information Commons: An Example.....	1614
b)	The Changing Bases of Knowledge and Rights in an Information Commons.....	1614
3.	<i>A Proposed Latency Analysis for Information Commons</i>	1616
a)	Knowledge Latency.....	1616
b)	Rights Latency.....	1618
c)	Latency Analysis	1618
d)	Implementing Latency Analysis in Commons Design	1620
III.	COMMONS IN THE SCIENCES	1621
A.	CONSTRUCTING SCIENCE COMMONS.....	1621
1.	<i>Incentives to Share</i>	1621
2.	<i>Modes of Data Sharing</i>	1625
3.	<i>Stakeholders and Policy Considerations</i>	1628

© 2011 Jorge L. Contreras.

[†] Acting Director of Intellectual Property Programs and Senior Lecturer in Law, Washington University in St. Louis. The author wishes to thank the following for their thoughtful comments: Scott Baker, Sara Bronin, Dan Burk, Rebecca Eisenberg, Terry Fisher, Charlotte Hess, Kimberly Kaphingst, Michael Madison, Charles McManis, Ruth Okediji, Nirej Sekhon, the participants in the 2009 SALT Junior Faculty Development Workshop, the 2010 Junior Scholars in Intellectual Property (JSIP) workshop and the 2010 Intellectual Property Scholars Conference (IPSC), and research assistant Jeff Johnson.

B.	EXOGENOUS INFLUENCES ON LATENCIES WITHIN A SCIENCE COMMONS	1630
1.	<i>Publication Delay</i>	1631
2.	<i>Copyright and Other Restraints on the Dissemination of Data</i>	1633
3.	<i>Patents and Restraints on the Use of Data</i>	1635
a)	Effect of Patents on Scientific Data.....	1635
b)	Patents and Latency	1636
c)	Patents and Anticommons.....	1637
IV.	TIMING AND DATA SHARING: TWO CASE STUDIES	1641
A.	THE GENOME COMMONS.....	1641
1.	<i>Development of Genomics Data Release Policies</i>	1641
a)	The Human Genome Project (HGP)	1641
b)	The Bermuda Principles	1644
c)	Bermuda, Data Release, and Patents	1645
d)	Growing Tension Between Data Generators and Users	1647
e)	GWAS and Embargo Policies	1648
f)	Private Sector Policies.....	1650
B.	SCIENTIFIC PUBLISHING.....	1652
1.	<i>Scientific Publishing and the Open-Access Debate</i>	1652
2.	<i>Open Access and Government-Funded Research</i>	1654
3.	<i>Alternative Open-Access Models</i>	1655
V.	APPLYING THE LATENCY FRAMEWORK	1657
A.	LATENCY IN THE GENOME COMMONS.....	1657
1.	<i>Stakeholders and Policy Design Considerations</i>	1657
2.	<i>Latency as a Policy Design Tool—Retention versus Embargo</i>	1660
B.	LATENCY AND SCIENTIFIC PUBLISHING.....	1664
1.	<i>Stakeholders and Policy Design Considerations</i>	1664
2.	<i>Compromises Effected Through Latency Choices</i>	1665
C.	SYNTHESIS AND BROADER APPLICATIONS.....	1667
VI.	CONCLUSION	1671

I. INTRODUCTION

The fossilized remains of *ardipithecus ramidus* (“*Ardi*”), an early human ancestor and important evolutionary link, were unearthed in 1992 by archeologists from the University of California, Berkeley, and the University of Tokyo. The scientists published their initial findings in 1994 to great

acclaim,¹ but did not release detailed data about the fossils until 2009,² seventeen years after the initial discovery. This long delay, justified by the researchers as necessary due to the fragile condition of the remains and the difficulty of extracting and reassembling them, has nevertheless been widely criticized as contrary to the practice of “open science.”³

Data is the currency of science, and the sharing of data is fundamental to the scientific enterprise.⁴ Since the emergence of modern scientific practice, scientists have published their observations and experimental data to earn professional recognition, permit verification of their theories, and further the overall progress of science.⁵ Today, powerful database and networking technologies have enabled the dissemination of data on a scale not

1. Tim D. White et al., *Australopithecus Ramidus, a New Species of Early Hominid from Aramis, Ethiopia*, 371 NATURE 306 (1994); see also John Noble Wilford, *Fossil Find May Show if Prehuman Walked*, N.Y. TIMES, Feb. 21, 1995, at C1 (calling *a. ramidus* “potentially the most significant discovery in early human studies since the Lucy fossils in 1974”).

2. Tim D. White et al., *Ardipithecus Ramidus and the Paleobiology of Early Hominids*, 326 SCI. 65 (2009).

3. See, e.g., Joel Achenbach, *Ancient Skeleton Could Rewrite the Book on Human Origins*, WASH. POST, Oct. 2, 2009 (noting “impatience” in the scientific community while waiting for the *Ardi* data); Editorial, *Fossils for All: Science Suffers by Hoarding*, SCI. AM., Aug. 2009, at 26 (“[F]ossil hunters often block other scientists from studying their treasures, fearing assessments that could scoop or disagree with their own. In so doing, they are taking the science out of paleoanthropology.”).

4. See JOHN M. ZIMAN, PROMETHEUS BOUND: SCIENCE IN A DYNAMIC STEADY STATE 40 (1994); Yochai Benkler, *Commons-Based Strategies and the Problems of Patents*, 305 SCI. 1110, 1110 (2004). In this context, I (and most other commentators) use the term “science” to refer to “basic” research, or the investigation of fundamental natural laws and properties, rather than applied research or technology development that is directed to commercial application. See, e.g., ZIMAN, *supra*, at 24–26; Richard R. Nelson, *The Simple Economics of Scientific Research*, 67 J. POL. ECON. 297, 300–01 (1959) (discussing the imprecise line between basic and applied research). The norms of openness that apply to basic research do not necessarily apply to these later categories of investigation, and many of these are, quite naturally, conducted by corporate and industrial concerns in secret. See, e.g., Arti Kaur Rai, *Regulating Scientific Research: Intellectual Property Rights and the Norms of Science*, 94 NW. U. L. REV. 77, 93 (1999).

5. See Robert K. Merton, *The Normative Structure of Science (1942)*, in THE SOCIOLOGY OF SCIENCE 267, 267–78 (Norman W. Storer ed., 1973); discussion *infra* Section III.A. Merton famously identifies four “norms” of science: universalism (scientific claims should be evaluated using objective criteria consistent with observation), communism (now generally referred to as “communalism”) (scientific findings should belong to the scientific community as a whole), disinterestedness (scientists should have no emotional or financial attachment to their work), and organized skepticism (scientists should act dispassionately, without regard to personal beliefs). See also ZIMAN, *supra* note 4, at 77; Benkler, *supra* note 4, at 1110 (“[O]pen distribution of the inputs and outputs of the scientific process are its organizational norms.”); Rai, *supra* note 4, at 90–91 (discussing in detail the norms of scientific research, including the Mertonian norm of communalism).

imaginable just a few decades ago. Developed primarily by government-funded research projects, vast collections of publicly accessible⁶ data now exist in fields such as chemistry,⁷ meteorology,⁸ geophysics,⁹ astronomy,¹⁰ paleontology,¹¹ and, as will be discussed *infra*, molecular biology and genomics.¹² Other data collections have been proposed in a wide range of additional disciplines.¹³ These aggregations of public scientific data, sometimes referred to as “science commons,”¹⁴ can serve as useful resources

6. My use of the term “publicly accessible” connotes that data may be accessed by any researcher, though not necessarily without charge. Some of the databases discussed in this Article charge for access, just as most scientific journals charge for subscriptions. For a discussion of this issue, see *infra* Section IV.B.1. Also, even to the extent that data is “publicly” available within a commons, various practical, technical, and logistical issues may impact individuals’ ability to access and use that data. See Jorge L. Contreras, *Prepublication Data Release, Latency and Genome Commons*, 329 SCI. 393, 394 n.16 (2010); see also *infra* note 63 and accompanying text.

7. See Robert Potenzzone, *Opportunities for Commercial Exploitation of Networked Science and Technology Public-Domain Information Resources*, in NAT’L RESEARCH COUNCIL, THE ROLE OF SCIENTIFIC AND TECHNICAL DATA AND INFORMATION IN THE PUBLIC DOMAIN: PROCEEDINGS OF A SYMPOSIUM 52, 53 (2003) [hereinafter NRC, PUBLIC DOMAIN].

8. See U.S. NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION’S (“NOAA”) NATIONAL DATA CENTERS, <http://nesdis.noaa.gov>; BRITISH ATMOSPHERIC DATA CENTER (BADc), <http://badc.nerc.ac.uk> (discussed in DIGITAL ARCHIVING CONSULTANCY ET AL., LARGE-SCALE DATA SHARING IN THE LIFE SCIENCES: DATA STANDARDS, INCENTIVES, BARRIERS AND FUNDING MODELS app. A-2 (2005) [hereinafter JOINT DATA STANDARDS STUDY]).

9. See U.S. GEOLOGICAL SURVEY’S EARTH RESOURCES OBSERVATION SYSTEMS (“EROS”) DATA CENTER, <http://eros.usgs.gov> (last visited Nov. 22, 2010).

10. See, e.g., M. Jordan Raddick & Alexander S. Szalay, *The Universe Online*, 329 SCI. 1028 (2010) (discussing the massive Sloan Digital Sky Survey (www.sdss.org)); NATIONAL AERONAUTICS AND SPACE ADMINISTRATION’S (NASA) SPACE SCIENCE DATA CENTER, <http://nssdc.gsfc.nasa.gov>; *About the PDS*, PLANETARY DATA SYSTEM (PDS), <http://pds.nasa.gov/about/about.shtml> (last visited Sept. 17, 2010).

11. See THE OPEN DINOSAUR PROJECT, <http://opendino.wordpress.com> (last visited Sept. 17, 2010).

12. See *infra* Section IV.A.

13. See, e.g., JAMES BOYLE, THE PUBLIC DOMAIN: ENCLOSING THE COMMONS OF THE MIND 171–78 (2008) (describing the MIT Registry of Standard Biological Parts); Geoff Brumfiel, *Chemists Spin a New Web of Data*, 453 NATURE 139 (2008) (describing a new open-access source of data on molecular chemistry); David Einhorn & Rita Heimes, Letter to the Editor, *Creating a Mouse Academic Research Commons*, 27 NATURE BIOTECHNOLOGY 890 (2009) (describing the impetus for creating a public resource for genetically modified mouse strains); Peter A. Stott & Peter W. Thorne, *How Best to Log Local Temperatures?*, 465 NATURE 158 (2010) (describing efforts to develop international data sharing of local temperature information).

14. This term finds its root in shared physical resources that have been broadly categorized as “commons.” See *infra* notes 19–23 and accompanying text; *infra* Section II.A. A “science commons” is a species of commons that is devoted to scientific data and

for the global scientific community and enable collaboration and information sharing at unprecedented levels.

But just as these public data resources proliferate, there are signs that all is not well in the science commons. As the example of *Ardi* strikingly illustrates, significant delays between the generation of scientific data and its disclosure to the public arise for a host of reasons, including publication lag time, intellectual property restrictions, and institutional and interpersonal inefficiencies.¹⁵ The recent outcry over BP's attempt to maintain the secrecy of data collected from the catastrophic *Deepwater Horizon* oil spill¹⁶ reflects growing public and academic unease with the concealment and delayed release of scientific data.¹⁷ This unease is shared by scholars who debate the benefits and effects of science-based commons, whether intellectual property protection threatens the existence of such commons, and whether intellectual property systems hinder or promote innovation and discovery.¹⁸

information. My use of the term "science commons" is not intended to reference the non-profit organization known as Science Commons, which undertakes various projects, including tool development, to make scientific data more accessible to the public. *See About Science Commons*, SCIENCE COMMONS, <http://sciencecommons.org/about/> (last visited Sept. 14, 2010).

15. *See infra* Section III.B.

16. *See* Amanda Mascarelli, *Freedom of Spill Research Threatened*, 466 NATURE 538 (2010); Lauren Schenkman, *After Outcry, Oil Data Inches into the Open*, 329 SCI. 888 (2010) (describing BP's retreat from its initial position following public criticism of its data concealment practices).

17. *See, e.g.*, ZIMAN, *supra* note 4, at 40–41 (“[Increasing secrecy] not only slows the advance of knowledge: it also puts a damper on public assessments of research claims, which are the ultimate arbiter of scientific validity.”); *see also* David Blumenthal et al., *Data Withholding in Genetics and the Other Life Sciences: Prevalences and Predictors*, 81 ACAD. MED. 137, 137 (2006) [hereinafter Blumenthal et al., *Prevalences and Predictors*] (“[D]ata withholding is common in biomedical science.”); David Blumenthal et al., *Withholding Research Results in Academic Life Science: Evidence from a National Survey of Faculty*, 277 J. AM. MED. ASSN. 1224 (1997) [hereinafter Blumenthal et al., *National Survey*] (reporting that 8.9% of academic life scientists have refused to share research results with other scientists within the past three years); Ryan P. O'Donnell et al., *Hindrance of Conservation Biology by Delays in the Submission of Manuscripts*, 24 CONSERVATION BIOLOGY 615 (2010) (examining delays in submission of data for publication in the field of conservation biology); Elizabeth Pisania et al., *Time for Fair Trade in Research Data*, 375 LANCET 703 (2010) (describing a lack of data sharing in the field of public health research); Josh Sommer, *The Delay in Sharing Research Data is Costing Lives*, 16 NATURE MED. 744 (2010); *Fight Over Access to 'Hobbit' Bones*, NEWSIDENTIST, Dec. 11, 2004, available at <http://www.newscientist.com/article/mg18424772.900-fight-over-access-to-hobbit-bones.html> (reporting that a researcher was isolating valuable samples from other researchers).

18. *See* BOYLE, *supra* note 13; Elinor Ostrom & Charlotte Hess, *A Framework for Analyzing the Knowledge Commons*, in UNDERSTANDING KNOWLEDGE AS A COMMONS: FROM

Commons theory offers an attractive and timely framework for the analysis of scientific data sharing. Over the past two decades, legal scholarship, economic analysis, and organizational theory have become deeply invested in the theoretical construct of the commons.¹⁹ Taking as its foundational metaphor the non-proprietary physical resources shared by traditional societies such as pastures, fisheries, and forests, the systematic study of commons systems was pioneered by Elinor Ostrom in the 1980s and early 1990s.²⁰ Her work in this area earned her the 2009 Nobel Prize in Economics. Among Ostrom's many insights was the application of the well-known Institutional Analysis and Development (IAD) framework applied by organizational theorists to privately ordered commons structures.²¹ Ostrom, together with Charlotte Hess, has recently extended this methodology to the analysis of so-called "information commons," shared intangible resources as varied as audiovisual content, open-source software, digital book archives, medical records, and traditional knowledge.²² More recently, Michael Madison, Brett Frischmann, and Katherine Strandburg have undertaken a thorough re-examination of the IAD framework in relation to commons in the "cultural environment."²³

Despite the abundance of debate and scholarship on this topic, until now, little attention has been paid to the *temporal* dimension of information commons—that is, the rate at which information is added to, or subtracted from, the commons.²⁴ These temporal characteristics are crucial, as they

THEORY TO PRACTICE (Charlotte Hess & Elinor Ostrom eds., 2006) [hereinafter KNOWLEDGE AS A COMMONS]; *infra* Section III.B.

19. A sense of the volume of this scholarship can be obtained from the online Digital Library of the Commons hosted at Indiana University, <http://dlc.dlib.indiana.edu/dlc/> (last visited July 24, 2010), which, at last count, included nearly 6,000 scholarly articles relating to the commons.

20. See ELINOR OSTROM, GOVERNING THE COMMONS — THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION (1990).

21. See *infra* Section II.A.2.

22. See Ostrom & Hess, *supra* note 18, at 42–43.

23. Michael J. Madison, Brett M. Frischmann & Katherine J. Strandburg, *Constructing Commons in the Cultural Environment*, 95 CORNELL L. REV. 657, 659 (2010). Madison, Frischmann, and Strandburg refer to aggregations of shared information as "cultural commons" and include within their far-ranging analysis shared resource structures as varied as patent pools, open source software, Wikipedia, the Associated Press, and jamband fan communities. *Id.* at 660–63. For purposes of this Article, I adopt Ostrom's terminology of "information commons," as this term more intuitively characterizes the scientific data that is the central focus of this Article.

24. It is important to distinguish between the dynamic nature of an information resource, which is the focus of this Article, and the dynamic nature of the community that interacts with and creates an information commons, which has been analyzed by others. For

define the very nature of the resource being shared and are essential to the overall functioning and value of the commons. In this Article, I describe a novel analytical tool, which I term “latency analysis,” for analyzing and designing the temporal features of information commons.²⁵ I place this analytical tool within the organizational frameworks proposed by Ostrom and Hess, and Madison, Frischmann, and Strandburg, though its applicability is not necessarily limited to these frameworks.

Latency analysis utilizes two key variables that characterize all information commons: the rate at which information enters the commons, its *knowledge latency*, and the rate at which the knowledge in the commons becomes freely utilizable, its *rights latency*.²⁶ With these two variables in mind, latency analysis provides a three-step analytical methodology that consists of (1) determining the stakeholder groups relevant to the information commons,²⁷ (2) determining the policy objectives that are relevant to each stakeholder group,²⁸ and (3) mediating among the differing positions of the stakeholder groups through adjustments in the latency variables of the commons.

The latency analysis that I develop in this Article is both descriptive and prescriptive. Not only is it a useful tool for analyzing existing information

example, while Ostrom and Hess acknowledge the dynamic nature of information commons, Ostrom & Hess, *supra* note 18, at 41–42 (“[T]he framework can also be used to analyze dynamic situations where individuals develop new norms, new rules, new physical technologies.”), the focus of their analysis is on the dynamics of the institutional and organizational *structures* that utilize the commons, rather than the dynamic characteristics of the commons themselves. Likewise, Madison, Frischmann, and Strandburg, *supra* note 23, at 682, acknowledge the “iterative” nature of rules and resources within a cultural commons and state that “[d]ynamic effects are central to [their] analysis.” *Id.* at 673. However, like Ostrom and Hess, they focus on the dynamic characteristics of the communities (musicians, open source developers, etc.) who participate in commons construction. While they refer briefly to the expiration of patents and copyrights when explaining that “[t]he durability of shared resources must be considered,” *id.* at 689, by and large they do not focus on the dynamic characteristics of the common resource itself.

25. I first described this approach in Contreras, *supra* note 6, at 393.

26. *Id.* at 393.

27. See Madison, Frischmann & Strandburg, *supra* note 23, at 690 (prescribing the identification of constituencies and their part in commons formation); Ostrom & Hess, *supra* note 18, at 48–50 (noting the importance of defining the “attributes of the community”).

28. See Madison, Frischmann & Strandburg, *supra* note 23, at 691–93 (identifying community goals and objectives); Ostrom & Hess, *supra* note 18, at 48–50 (considering goals and objectives among the attributes of the community).

commons;²⁹ it also offers policy designers an objective set of variables with which to mediate otherwise value-based negotiations among stakeholders.³⁰ The introduction of latency variables into the commons formation discourse thus enables parties to avoid often fruitless debate over entrenched values-based preferences (for example, intellectual property protection over the public domain) and to achieve compromise based on simple numerical calculations.³¹ In this sense, the use of latency variables in policy design helps to reduce the transaction costs of negotiating commons policy, to achieve efficient and equitable results for all stakeholders, and thereby to facilitate the formation of socially valuable commons of information.

In Part II, I review the history and current state of commons theory, then describe latency analysis and its place within existing analytical frameworks for the commons. In Part III, I describe the formation of commons in the sciences, relevant stakeholder groups, and modes of data sharing. I then address the principal exogenous factors that have been implicated in limiting the growth of science commons, particularly delays induced by the scientific publication process and intellectual property rights that either limit the entry of knowledge into a commons (i.e., copyright and trade secret) or limit the usability of knowledge once it enters (i.e., patent). In this Part, I also discuss the well-known debate over the “anticommons” and its implications for the formation and utility of science commons.³²

In Part IV, I describe two well-known narratives of commons formation in the sciences: the “genome commons”³³ and open-access scientific publishing.³⁴ In each of these cases, I analyze the cultural, legal, political, and historical influences that shaped the development of the relevant scientific commons, as well as the competing policy objectives of the relevant

29. In this sense, answering the call of Madison, Frischmann, and Strandburg for greater study of existing commons structures and institutions. Madison, Frischmann & Strandburg, *supra* note 23, at 708–09.

30. *See infra* Section V.C.

31. This technique has long been recognized as an effective negotiation practice. *See, e.g.*, ROGER FISHER, WILLIAM URY & BRUCE PATTON, GETTING TO YES: NEGOTIATING AGREEMENT WITHOUT GIVING IN 4–12, 81–92 (2d ed. 1991) (illustrating the differences between “positional” and “principled” negotiation and the value of negotiating from objective criteria).

32. *See* Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCI. 698 (1998).

33. I adopt this term in Contreras, *supra* note 6.

34. Madison, Frischmann, and Strandburg advocate blending the traditional functionalist account of cultural commons with a “metaphorical or narrative approach.” Madison, Frischmann & Strandburg, *supra* note 23, at 671.

stakeholder groups. Section IV.A begins with an overview of the groundbreaking set of privately ordered principles³⁵ adopted in 1996 requiring that all genetic sequence data generated by the Human Genome Project (HGP) be released in publicly accessible databases a mere twenty-four hours after generation. The Bermuda Principles continue to shape data release practices in genomics and other fields,³⁶ but not without some important caveats. Most significant among these are timing-based policies that were implemented to mediate among the competing interests of data-generating scientists, data-using scientists, and public funding institutions, and which have resulted in two distinctly different approaches to commons formation. In Section IV.B, I discuss the recent debate over “open access” scientific publishing and the commons of scientific knowledge more broadly. In particular, I draw attention to private bilateral compromises among research institutions and publishers as well as new federal policies that are intended to balance the competing interests of publishers and the public by imposing specified time periods before the public release of scientific papers is required.

In Part V, I apply the methods of latency analysis to each of these case studies, analyzing in greater detail the various timing-based policy features that enabled compromise among competing stakeholder positions. I then draw a number of conclusions regarding the success of latency-based mediation of commons policy negotiation positions and suggest areas in which further theoretical and empirical work may be of value.

II. COMMONS AND LATENCY

35. HUGO, *Summary of Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing*, HUMAN GENOME PROJECT INFORMATION (1996), http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml/ [hereinafter Bermuda Principles]. A more detailed discussion of the Bermuda Principles can be found in Jorge L. Contreras, *Bermuda's Legacy: Policy, Patents and the Design of the Genome Commons*, 12 MINN. J. L. SCI. & TECH. 61 (2011).

36. See, e.g., Francis Collins, *Opinion: Has the Revolution Arrived?*, 464 NATURE 674, 675 (2010) (referring to the “radical ethic of immediate data deposit” adopted by the HGP as the current “norm for other community research projects”); Jane Kaye et al., *Data Sharing in Genomics — Re-shaping Scientific Practice*, 10 NATURE REV. GENETICS 331, 332 (2009) (“[T]hese policies have created a climate in which data sharing has become the default, and [grant] applicants must demonstrate why their data should be exempt from the requirement that it should be deposited for use by other scientists.”); Nikos Kyrpides, *Fifteen Years of Microbial Genomics: Meeting the Challenges and Fulfilling the Dream*, 27 NATURE BIOTECHNOLOGY 627, 627 (2009) (“Over time, as the substantial benefits of prepublication release of genome data have been recognized, many funding agencies and most of the large sequencing centers now adhere to the rapid data release policy set forth as the Bermuda Principles.”).

A. COMMONS THEORY BACKGROUND

1. *Physical Resource Commons*

Since the Middle Ages, the term “commons” has denoted shared physical spaces such as fields, pastures, and forests that were open and free for exploitation by farmers, herdsmen, and other local peoples.³⁷ In 1968, biologist Garrett Hardin described an effect he termed the “tragedy of the commons,” rooted in Malthusian notions of population growth and the deleterious effects of unbounded utilization of depletable common resources.³⁸ Hardin employed the metaphor of a shared pasture to illustrate the social harm that may arise when individual herdsmen seek to maximize their own gain by allowing ever more animals to graze on the common land. The result, of course, is the critical depletion or destruction of the shared resource. “Therein is the tragedy,” Hardin wrote, “[e]ach man is locked into a system that compels him to increase his herd without limit—in a world that is limited.”³⁹ For Hardin, a “commons” connoted a resource that was accessible to many without constraint, such as a pasture, a forest, or the ocean. Later scholars have defined this attribute of commons as “non-excludability”: the inability to limit the use of a resource due to its inherent features, such as size or accessibility.⁴⁰

37. See Ostrom & Hess, *Introduction to KNOWLEDGE AS A COMMONS*, *supra* note 18, at 12; NANCY KRANICH, *THE INFORMATION COMMONS — A PUBLIC POLICY REPORT 10* (2004), available at <http://www.fepproject.org/policyreports/InformationCommons.pdf>. In the U.S., “commons” have also been associated historically with New England’s open town squares that served as popular venues for speechifying and pamphleteering. Ostrom & Hess, *supra*, at 13. In both cases, “commons” terminology has a strong traditional association with freedom and openness.

38. Garrett Hardin, *The Tragedy of the Commons*, 162 *SCI.* 1243, 1244 (1968). The first modern economic analyses of commons structures are usually credited to H. Scott Gordon and Anthony Scott, who based their models on studies of fisheries in the mid 1950s. See Charlotte Hess & Elinor Ostrom, *Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource*, 66 *LAW & CONTEMP. PROBS.* 111, 115–16 (2003).

39. Hardin, *supra* note 38, at 1244. Hardin’s concern is not specifically with pasture land, but with all scenarios in which natural resources may be depleted due to a “tragedy of the commons,” including ocean fish stocks, national parks, the environment, and the earth’s ability to support a rapidly growing population. Elinor Ostrom cites numerous additional problems that have been analyzed using commons theory, including famines, acid rain, and urban crime. OSTROM, *supra* note 20, at xv.

40. See OSTROM, *supra* note 20, at 30. Other definitions and variations on this theme have been addressed in the literature. See, e.g., Hess & Ostrom, *supra* note 38, at 114–18 (summarizing several commons approaches and theories).

As noted above, Elinor Ostrom conducted the seminal analysis of social and organizational structures governing physical commons in the 1980s.⁴¹ Among Ostrom's many insights was the applicability of the well-known Institutional Analysis and Development (IAD) framework, employed since the 1970s to evaluate organizational characteristics and institutional decision-making, to common-pool resources.⁴² Under the IAD framework, commons structures may be examined with respect to three broad sets of characteristics: (1) those of the common resource itself, (2) the "action arena" in which the common resource is utilized, and (3) the desired or actual outcomes of the commons structure.⁴³ Each of these broad areas may be subdivided into further analytical components, so that the common resource, for example, is assessed with respect to its bio-physical characteristics, the attributes of the relevant community, and the applicable rules set, whether legal or norms-based.⁴⁴ The application of the IAD framework analysis results in a deeper understanding of the factors that should be considered when structuring or evaluating a commons structure, and Ostrom and others have persuasively applied the IAD framework to common resource arrangements ranging from fisheries to irrigation systems to environmental governance.⁴⁵

2. *Commons of Information*

In the mid-1990s, scholars began to apply commons theory to intangible shared resources and information.⁴⁶ Since then, much has been written about so-called "information commons" in areas including computer software, network capacity, artistic content, scholarly learning, and scientific data.⁴⁷ In

41. OSTROM, *supra* note 20.

42. See Charlotte Hess & Elinor Ostrom, *A Framework for Analysing the Microbiological Commons*, 58 INT'L SOC. SCI. J. 335, 339 (2006); Ostrom & Hess, *supra* note 18, at 42–43.

43. See Ostrom & Hess, *supra* note 18, at 44–45.

44. *Id.* at 45–53.

45. See Hess & Ostrom, *supra* note 42, at 339.

46. Hess & Ostrom, *supra* note 37, at 4 (noting the "explosion" of information commons scholarship beginning around 1995).

47. See, e.g., HAL ABELSON, KEN LEDEEN & HARRY LEWIS, BLOWN TO BITS — YOUR LIFE, LIBERTY, AND HAPPINESS AFTER THE DIGITAL EXPLOSION 277 (2008) (discussing the application of commons theory to broadcast spectrum); LAWRENCE LESSIG, THE FUTURE OF IDEAS 85–86 (2001) (arguing that commons systems have encouraged innovation, specifically with respect to software, telecommunications, and the Internet); JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET AND HOW TO STOP IT 78–79 (2008) (discussing commons approaches both to Internet content and hardware); Yochai Benkler, *Coase's Penguin, or Linux and the Nature of the Firm*, 112 YALE L.J. 369 (2002) (arguing that "commons-based peer production" of software has proven to be both viable and efficient, as

these discussions, information and technology are analogized to the grasses in Hardin's pasture, as public knowledge is, by its nature, non-excludable. That is, once an item of information becomes generally known, it is difficult (absent the contractual or statutory structures described below) to prevent others from sharing it.⁴⁸

Information commons are, of course, different than aggregations of finite physical resources inasmuch as information is generally viewed as "non-rivalrous," meaning that any number of individuals may enjoy its benefits without depleting it: a fact, whether known by one person or by one million, remains undiminished.⁴⁹ As such, Hardin's "tragedy of the commons," which arises from self-interested over-exploitation of a common resource, is unlikely to occur within the context of information commons. However, as will be discussed below, other potential "tragedies" that are more specific to information commons have been postulated.⁵⁰

Building upon their earlier work on physical commons, Ostrom and Hess have applied the IAD framework to the analysis of knowledge-based commons structures, reasoning that both physical resource commons and information commons share numerous attributes.⁵¹ They caution, however, that further evolution of the IAD model may be required to understand the peculiar attributes of information commons more thoroughly.⁵² Michael Madison, Brett Frischmann, and Katherine Strandburg recently accepted this challenge and have undertaken a thorough re-examination of the IAD

demonstrated by the model of the Linux operating system); James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 LAW & CONTEMP. PROBS. 33, 44–49 (2003) (discussing open source software).

48. See, e.g., Hess & Ostrom, *supra* note 37, at 8–9. In much of the political and economics literature on commons, a distinction is drawn between "public goods" such as sunsets, which are owned by no one but which may be enjoyed by all and "common pool resources," such as highways and public libraries, which may be owned by a government or collective, but which are open to enjoyment by all. See Hess & Ostrom, *supra* note 38, at 119–21. In these terms, this Article addresses common-pool resources as opposed to naturally occurring public goods.

49. See, e.g., Dana Dalrymple, *Scientific Knowledge as a Global Public Good: Contributions to Innovation and the Economy*, in NRC, PUBLIC DOMAIN, *supra* note 7, at 35, 48 (noting that scientific knowledge, in its pure form, is "the epitome of a global public good," as it is "freely available to all and is not diminished by use—indeed it may grow with use"); Hess & Ostrom, *supra* note 37, at 8–9; Boyle, *supra* note 47, at 41–42. But see David W. Opperbeck, *The Penguin's Genome, or Coase and Open Source Biotechnology*, 18 HARV. J.L. & TECH. 167, 209–10 (2004) (seeking to refute the characterization of information as non-rivalrous).

50. See *infra* Section III.B and accompanying notes.

51. Ostrom & Hess, *supra* note 18, at 43.

52. *Id.* at 68.

framework in relation to commons in the “cultural environment.”⁵³ In doing so, they recognized that, unlike the farmers and fishermen who exploit physical commons of natural resources, users of information commons not only *use* the common resource, but *produce* it as well.⁵⁴ This insight led them to propose a modified framework that more closely links the features of the common resource to its user–producers, as mediated through constructed “rules in use,” and which also seeks to combine the functionalist approach of the IAD framework with metaphorical and narrative accounts of commons formation.⁵⁵

B. THE TEMPORAL DIMENSION OF INFORMATION COMMONS

1. *Dynamic and Static Commons*

The classical physical common resource posited by Hardin and others is both shared and finite, meaning that, absent husbandry and control, the resource will be over-consumed and therefore decline and possibly disappear over time: if the pasture continues to be over-grazed, it will become depleted within a predictable timeframe.⁵⁶ Curiously, the discussion of information commons typically lacks this dynamic element. Whereas physical commons are often characterized specifically in terms of their alteration (usually diminution) over time, information commons are more frequently described in one state or another without much reference to dynamic alteration of the resource, and *quantitative* changes to the commons have not traditionally been viewed as analytically significant, except in the most general terms.⁵⁷ Certainly, close analysis of information commons in their static state is a necessary and important exercise, as the fundamental arguments regarding, for example, the effects of intellectual property protection on common resources can be considered time-independent. Moreover, information commons are different from physical resource commons in that the contents of information commons generally do not diminish with increased use and

53. Madison, Frischmann & Strandburg, *supra* note 23, at 659.

54. *Id.* at 681. In this respect, they echo the well-known principle that users of intellectual property are also its producers. See WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* 13–14 (2003).

55. Madison, Frischmann & Strandburg, *supra* note 23, at 671, 681–83.

56. This is Hardin’s “tragedy of the commons.” See *supra* notes 38–39 and accompanying text. For a general discussion of the temporal depletion of physical resource commons, see Ostrom’s classic work *GOVERNING THE COMMONS*, *supra* note 20.

57. See, e.g., David Bollier, *Growth of the Commons Paradigm*, in *KNOWLEDGE AS A COMMONS*, *supra* note 18, at 27, 34 (“[M]any information commons exemplify what some commentators have called ‘the cornucopia of the commons,’ in which more value is created as more people use the resource . . . the operative principle is ‘the more the merrier.’”).

their variance over time could therefore be considered less noteworthy than that of physical resource commons. Nevertheless, as will be discussed in the next Section, information commons *do* change over time, and the analysis of these temporal changes is important to understanding and designing information commons.

2. *Knowledge, Rights, and Timing in Information Commons*

a) The Dynamic Character of Information Commons: An Example

Information commons, by their nature, are dynamic. Take, for example, the body of knowledge concerning Jupiter's moons. Before Galileo discovered Jupiter's four largest satellites in 1610, the accumulated body of knowledge concerning them was zero. No information at all existed, and thus there could be no information commons. When the noted astronomer published his discovery a few months after it was made, an item of public knowledge was created and the commons relating to Jupiter's moons was born. As telescopes improved and astronomers' understanding of the solar system advanced over the next four hundred years, this information commons expanded. Recently, the aptly named Galileo spacecraft has provided detailed images and measurements of the Jovian satellites and their atmospheres, resulting today in a body of public information concerning the moons of Jupiter that is quite large.⁵⁸

b) The Changing Bases of Knowledge and Rights in an Information Commons

The above example relating to Jupiter's moons is generalizable to most information commons: the pool of data constituting the commons may expand and contract over time.⁵⁹ For purposes of discussion, I will term the total pool of data constituting an information commons at any given time its

58. For a concise overview, see Joseph A. Burns, *The Four Hundred Years of Planetary Science Since Galileo and Kepler*, 466 NATURE 575 (2010).

59. Some information commons, of course, *are* relatively static. Take, for example, a compilation of the works of Shakespeare. This body of work, though subject to periodic minor additions and subtractions, and notwithstanding the occasional effort (by no less august personages than the justices of the U.S. Supreme Court, see Jess Bravin, *Justice Stevens Renders an Opinion on Who Wrote Shakespeare's Plays*, WALL ST. J., Apr. 18, 2009, at 1) to claim or refute the Bard's authorship of a particular work, has remained more or less constant for more than a hundred years. The same is less true of science commons, which, in nearly all fields, tend to change as new experiments are conducted, observations are made, and hypotheses are advanced, withdrawn, refuted, and validated.

“knowledge base.”⁶⁰ The size of the knowledge base and the rate at which it changes are related variables by which an information commons can be measured and evaluated.

Just as the pool of *data* within an information commons may expand and contract over time, so may the set of *rights* applicable to the information within the commons. That is, for a given commons, the nature and duration of the usage restrictions on each data element may evolve over time, and the aggregate pool of usable data within the commons will likewise change.⁶¹ For purposes of this discussion, I will term the portion of the knowledge base of an information commons, the use of which is materially encumbered,⁶² as its “encumbered knowledge base.” Similarly, the portion of its knowledge base,

60. It is not uncommon to refer to this total pool of data as the information commons itself. However, for purposes of this analysis, it is helpful to distinguish between (a) the information commons, which includes its knowledge base, its “exogenous” characteristics, and the rules and policies governing its constitution, and (b) the size of the knowledge base associated with this commons at distinct periods in time.

61. There is a tendency in the discourse of commons to equate, or at least strongly associate, information commons with the public domain. *See, e.g.*, Anupam Chander & Madhavi Sunder, *The Romance of the Public Domain*, 92 CALIF. L. REV. 1331, 1338 (2004) (explicitly equating “commons” with the public domain, both generally and specifically in the context of indigenous genetic resources and traditional knowledge). This linkage has some appeal, as enjoying the benefits of a traditional commons (e.g., an open pasture) can be characterized by having the liberty to avail oneself fully of the benefits afforded by the resources within the “commons” (e.g., allowing one’s cattle to consume as much grass as they wish). The analogy between traditional commons and information commons, however, is imperfect. Unlike eating grass, catching fish, or timbering a forest, the “consumption” of resources in an information commons may occur in many different modes. For example, a digital music file in a “music commons” may be listened to, it may be copied, it may be transmitted, it may be mixed with other works, it may be synchronized with video footage, and it may be publicly performed, to name just a few potential modes of consumption. Each of these uses implicates distinct rights of the rights owner, and it is not necessary that a user be permitted to exercise every one of these rights in order to consider the music file part of a “commons.” Thus, for purposes of this Article, I proceed on the basis that an “information commons” may include knowledge that is covered by intellectual property rights, the use of which is thereby encumbered. *See* Boyle, *supra* note 47, at 68 (recognizing multiple “public domains”); Hess & Ostrom, *supra* note 38, at 121–22 (addressing the confusion between common-property and open-access regimes); J.H. Reichman & Paul F. Uhler, *A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment*, 66 LAW & CONTEMP. PROBS. 315, at 318–19 n.10 (2003) (distinguishing between “public domain” and “open access”). Madison, Frischmann, and Strandburg refer to structures including both public domain and encumbered knowledge as “semi-commons.” Madison, Frischmann & Strandburg, *supra* note 23, at 668.

62. By “materially encumbered” I mean that one or more material restrictions on the use of the data exist. These might include a contractual or policy embargo on presentation or publication of further results based on that data. At the extreme end of the temporal spectrum, patent rights can be viewed as another variety of encumbrance.

the use of which is not materially encumbered, or which is generally accessible and usable by a relevant public community,⁶³ as its “unencumbered knowledge base.”

Accordingly, each information commons may be characterized both in terms of its overall knowledge base and its unencumbered knowledge base (together with its complementary encumbered knowledge base). The size of an information commons’ overall knowledge base may or may not be closely correlated with the size of its unencumbered knowledge base. Thus, there may be large bodies of publicly accessible information (large knowledge base) that are almost entirely encumbered (small unencumbered knowledge base). One such example would be the database of knowledge contained in issued U.S. patents. It contains much information, but little freedom to use it, at least in the near-term.⁶⁴ Other information commons, such as that relating to the moons of Jupiter, may contain less information, but may have very few limitations on use. Thus, each information commons has varying complexions of knowledge and rights. And, by extension, as the character of an information commons changes over time, the relationship between these two variables—knowledge and rights—also changes, providing the basis for the latency analysis presented in the next Section.

3. *A Proposed Latency Analysis for Information Commons*

a) Knowledge Latency

For any given data element that is intended for inclusion in an information commons, there will be a period of time between its creation and its entry into the commons, when it becomes accessible to the relevant community. I term this time period “knowledge latency,” signifying that there is a delay, ranging from quite short to quite long, before which a particular data element destined for the commons actually appears there.⁶⁵

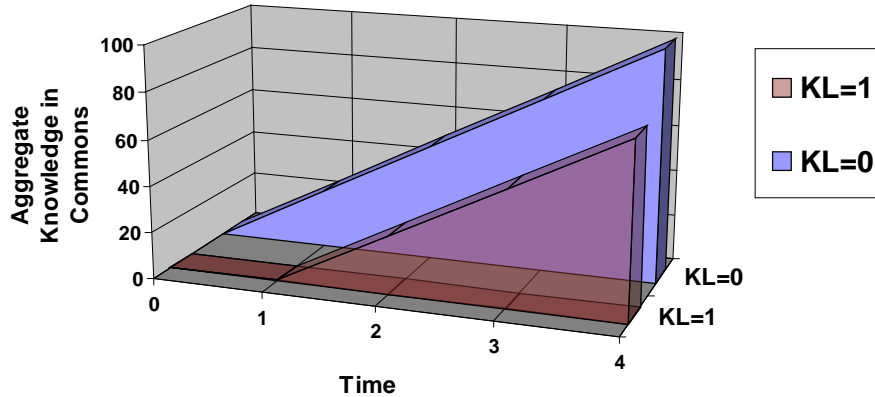
63. *See supra* note 6. Just as a traditional commons, such as a pasture or a forest, need not be accessible to the entire world, but only to the farmers and residents in the immediate area, I do not impose a requirement that information commons be accessible to the entire world. Rather, so long as the relevant community, be it of scientists, policy makers, or other interested users, has access to the commons, I view this as sufficient to qualify the body of information as accessible.

64. I have deliberately limited this example to “issued” patents. Of course, following expiration, the information contained in a patent is freely usable, as discussed *infra* notes 141–43 and accompanying text. I also assume, for purposes of this illustration, that the majority of inventive disclosures contained within a patent are protected by the patent claims and thus may not be used without infringement of these claims.

65. In information commons governed by policies with data deposit/release requirements, the maximum knowledge latency for the commons may be expressed as a

Knowledge latency therefore is an indicator of the rate of growth of the commons⁶⁶ and the amount of information available within the commons at any given time (Figure 1).

Figure 1: Effect of Knowledge Latency (KL) in an Information Commons⁶⁷



Knowledge latency in a given information commons may be expressed either as a *mandated* value derived from “rules in use” of the commons, or as an *actual* value. The *actual* value for knowledge latency may deviate from the *mandated* value for a number of reasons, including technical variations in data-deposit practices and intentional or inadvertent non-compliance by data generators. As with any set of policy-imposed timing requirements (e.g., time periods for making filings with governmental agencies), it is important to consider the mandated time delay for the deposit of data to an information commons. Because a mandated value is also, theoretically, the maximum amount of time that should elapse before a data element is deposited in the

single numerical value. In commons having data deposit requirements that are not quantified, such as those requiring that data be deposited as rapidly as possible following verification or based on other non-numerical criteria, knowledge latency may be expressed as a value range or estimated value, based on reasonable expectations and practice in the relevant field.

66. Note that knowledge latency shifts (delays) the time at which knowledge enters the commons, but not the *rate* at which knowledge enters the commons.

67. Figure 1 illustrates the aggregate Knowledge Base of an information commons over time, as determined by its Knowledge Latency. In the commons with KL=0, the Knowledge Base begins to grow at Time=0, whereas in the commons with KL=1, the Knowledge Base does not begin to grow until Time=1, resulting in a smaller Knowledge base at each subsequent time point.

commons, I generally refer to knowledge latency in terms of its *maximum* value.⁶⁸

b) Rights Latency

Just as there may be a delay between the generation of a data element and its deposit into an information commons, there may also be a delay between the appearance of data in the commons and its free usability, that is, its entry into the unencumbered knowledge base. I term this delay “rights latency.” As with knowledge latency, the term may be applied to an individual data element (i.e., representing the time before a particular data element becomes freely usable) or to the commons as a whole (i.e., representing the *maximum* time that it will take for data within the commons to become freely usable). The rights latency for a particular information commons may reflect a variety of factors, including policy-imposed embargos on the use of data and, in the extreme case, patent rights.⁶⁹ True public domain commons such as a compendium of Shakespeare’s works, in which no copyright or contractual encumbrances exist, would have rights latencies of zero. Commons that include data covered by patents would have rights latencies equal to the remaining patent term. Most information commons would fall somewhere between these two extremes.

c) Latency Analysis

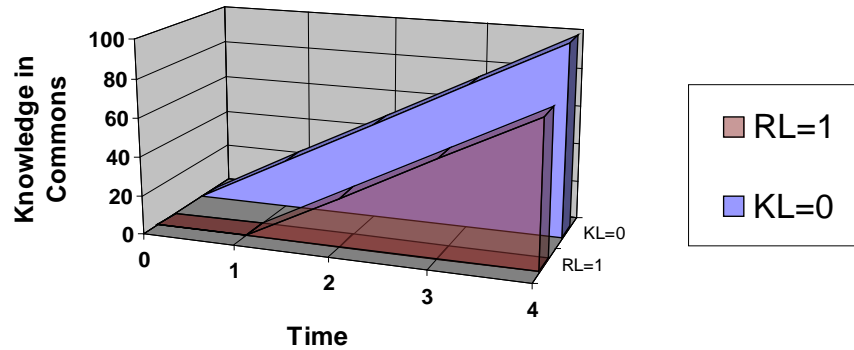
The related variables knowledge latency and rights latency describe important and complementary aspects of an information commons. If, as illustrated in Figure 2, rights latency is other than zero, the quantity of usable information in the commons will lag behind the total knowledge base.⁷⁰

68. Actual or observed knowledge latency values (both in terms of median values and the range of observed values) may be useful, however, in retrospective evaluation of user compliance with policy requirements or in assessing the efficiency of commons systems, and future analyses may find it useful to approach knowledge latency in this manner.

69. It may seem counter-intuitive to treat contractual limitations on use in a manner comparable to patent-based limitations on commercial exploitation. However, both of these “restrictions” serve to remove an item of information from public use. *See, e.g.*, Boyle, *supra* note 47, at 37. In each case, a user of the commons is aware of the restricted information and the knowledge that he or she obtains adds to his or her general store of knowledge, informs his or her view of the field, and apprises him or her of the developments and work of the data generator. In each case, however, the data user is unable to derive a direct benefit from the encumbered data, be that benefit a presentation or publication, or the development of a commercial product.

70. For purposes of this discussion, I assume that once established for a given information commons, knowledge latency and rights latency do not change substantially. This assumption, however, will not always hold true: commons “rules in use” may be

Figure 2: Effect of Knowledge Latency (KL) and Rights Latency (RL) in an Information Commons

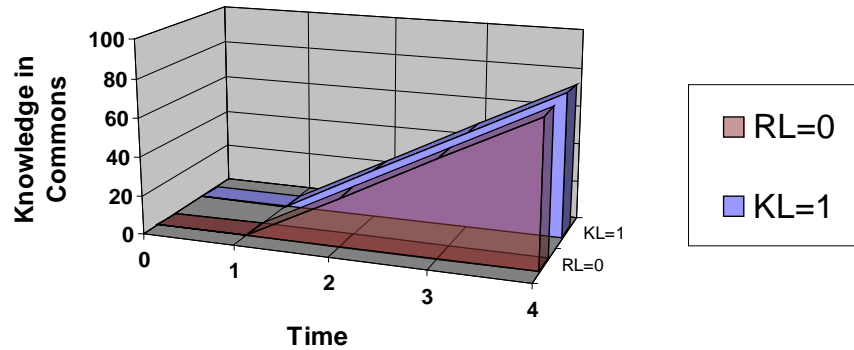


In the example illustrated by Figure 2, the rules of the commons establish that information be deposited into the commons as soon as it is created (KL=0), but a restriction, contractual or otherwise, limits its use until the passage of one time interval (RL=1). The line KL=0 shows the hypothetical increase of total knowledge within the commons beginning at the initial time (as there is no delay due to knowledge latency), and the area beneath line KL=0 represents the total knowledge within the commons. The line RL=1 shows the increase of unencumbered knowledge within the commons, and the area beneath the line RL=1 shows the total unencumbered knowledge within the commons.

An alternative scenario is shown in Figure 3. In this example, knowledge is withheld from the commons for one time interval after it is created (KL=1), but is released to the commons without encumbrance on future use (RL=0). The lines KL=1 and RL=0 are overlaid, resulting in a quantity of *unencumbered* knowledge that is equivalent to that illustrated in Figure 1, but less overall knowledge at any given time than the commons in Figure 2. The practical implications of these different design choices are illustrated by the case study in Section IV.A.

amended following the initial commons formation, and changes in law and policy may also impact the knowledge latency and rights latency of an information commons.

Figure 3: Effect of Knowledge Latency (KL) and Rights Latency (RL) in an Information Commons



d) Implementing Latency Analysis in Commons Design

As demonstrated by the above examples, latency variables are important elements in the analysis of information commons. However, the importance of these variables goes beyond their mere descriptive value. As discussed in Part V, the adjustment of rules that govern the latency characteristics of an information commons can enable policy designers to achieve varying policy goals. Moreover, the adjustment of latency variables can effectively mediate between the requirements of competing stakeholder interests and enable the creation of commons where disagreement might otherwise have precluded it.

In this Article, I propose a straightforward methodological approach for applying latency analysis to the evaluation and design of information commons. This approach consists of three steps which may be integrated into the frameworks of Ostrom and Hess,⁷¹ and Madison, Frischmann, and Strandburg⁷²: (1) determine the relevant communities of stakeholders of the information commons and their characteristics,⁷³ (2) determine the policy considerations relevant to each stakeholder group and the initial positions of each such group,⁷⁴ and (3) adjust the latency variables of the commons

71. See Ostrom & Hess, *supra* note 18, at 48–50.

72. Under both formulations, “rules in use” dictating these latencies will have an effect on the nature of the common resource. See Madison, Frischmann & Strandburg, *supra* note 23, at 680–82.

73. See Ostrom & Hess, *supra* note 18, at 48–50 (defining the “attributes of the community”); Madison, Frischmann & Strandburg, *supra* note 23, at 690 (prescribing the identification of constituencies and their part in commons formation).

74. See Ostrom & Hess, *supra* note 18, at 48–50 (considering goals and objectives among the attributes of the community); Madison, Frischmann & Strandburg, *supra* note 23, at 691–93 (identifying community goals and objectives).

(knowledge latency and rights latency) to mediate among the differing positions of the stakeholder groups.

III. COMMONS IN THE SCIENCES

The corpus of public scientific information has long been viewed as the quintessential information commons.⁷⁵ In this Part, I describe the norms and practices that have shaped the development of science commons, identify the various stakeholders invested in the development and use of such commons, and address some of the challenges that affect the growth of science commons today. In particular, I focus on three sets of exogenous factors that impact the addition of useful information to the commons: publication delay, copyright and other access-limiting rules, and patent-based restrictions on usage of information in the commons. The factors described in this Part form the basis of the policy objectives and stakeholder goals that will be addressed in Part IV. These competing objectives can effectively be mediated through the adjustment of latency variables.

A. CONSTRUCTING SCIENCE COMMONS

1. *Incentives to Share*

In the 1940s, sociologist Robert K. Merton famously identified the willingness of scientists to share their findings and experimental data as one of the fundamental norms that characterize both the practice and culture of science.⁷⁶ This norm finds its origin, however, far earlier than Merton's day; it is often traced to the late sixteenth and early seventeenth centuries—the age of Galileo, Bacon, and Newton. Prior to this time, specialized technical knowledge was typically guarded by trade guilds and individual practitioners who had little incentive to share with outsiders.⁷⁷ But beginning during what has traditionally been termed the “scientific revolution,”⁷⁸ new norms among scientists emerged that favored the open publication of ideas and the sharing of data. The causal factors that led to the adoption and spread of these “open science” norms are subject to debate among historians of science and are

75. See Benkler, *supra* note 4, at 1110.

76. See *supra* note 2 and accompanying text.

77. See Robert P. Merges, From Medieval Guilds to Open Source Software: Informal Norms, Appropriability Institutions, and Innovation 6–7 (Nov. 13, 2004), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=661543.

78. This term has been criticized by historians of science who contend that the emergence of modern science arose from a gradual process rather than a sudden “revolution.” See PETER J. BOWLER & IWAN RHYS MORUS, MAKING MODERN SCIENCE: A HISTORICAL SURVEY 23–25 (2005).

beyond the scope of this Article.⁷⁹ Nevertheless, the legacy of these significant social changes continues to influence the patterns and customs of scientific practice today.

Merton rationalized the sharing of scientific data on three functional bases. First, the advancement of science depends upon scientists' access to, and use of, previously collected data.⁸⁰ Little progress would be made if individual scientists were required to reproduce all the experiments of their predecessors. Thus, in order for science to advance, scientists must build upon earlier results in a cumulative fashion, resulting in an ever-expanding body of scientific knowledge that is available and accessible to the scientific community at large.⁸¹ Second, scientific data must be shared to enable scientists to validate and independently verify the analyses and conclusions of their colleagues.⁸² Recent instances of scientific fraud and the perception that misconduct among scientists is increasing underscore the need for critical

79. Paul David attributes the rise of "open science" to the intersection of late-Renaissance aristocratic patronage with the new class of experimental philosophers. Paul David, *Common Agency Contracting and the Emergence of "Open Science" Institutions*, 88 AM. ECON. REV. 15 (1998). Peter Bowler and Iwan Rhys Morus associate the emergence of modern scientific institutions and attitudes with the rise of universities. BOWLER & MORUS, *supra* note 78, at 322–26. Adrian Johns points to the success of the first scientific journal, the *Philosophical Transactions*, as indicative of the new style of collaborative experimental philosophy. ADRIAN JOHNS, *PIRACY: THE INTELLECTUAL PROPERTY WARS FROM GUTENBERG TO GATES* 62–63 (2009). And Elizabeth Eisenstein credits the printing press with the rise of scientific data sharing. ELIZABETH L. EISENSTEIN, *THE PRINTING PRESS AS AN AGENT OF CHANGE* 520–21 (1979).

80. Merton, *supra* note 5, at 274–75.

81. In this respect, Merton cites the well-known remark attributed to Sir Isaac Newton, "If I have seen farther it is by standing on the shoulders of giants." Merton, *supra* note 5, at 274–75. Merton's sentiment is echoed by contemporary analysts. See, e.g., JOINT DATA STANDARDS STUDY, *supra* note 8, at 11 ("[D]ata sharing contributes to a virtuous circle, where promoting effective sharing widens research and enhances scientific impact."). Today, the positive effects of data sharing are manifested both in the elimination of duplicative experiments and in the creation of large data sets that could not be created by individual scientists or groups. See Mike May, *Sharing the Wealth of Data*, SCIENTIFIC AMERICAN — WORLDVIEW, 88, 89–90 (2009).

82. See also NAT'L ACAD. OF SCIS., *ENSURING THE INTEGRITY, ACCESSIBILITY, AND STEWARDSHIP OF RESEARCH DATA IN THE DIGITAL AGE* 55 (2009) [hereinafter NAS, RESEARCH DATA] ("Only when a researcher shares data and results with other researchers can the accuracy of the data, analyses, and conclusions be verified."); Paul David, *The Economic Logic of "Open Science" and the Balance Between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer*, in NRC, *PUBLIC DOMAIN*, *supra* note 7, at 21 ("Disclosure . . . creates an expectation that all claims to have contributed to the stock of reliable knowledge will be subjected to trials of verification, without insult to the claimant.").

and independent review of scientific claims.⁸³ Similar considerations have also contributed to calls for greater public access to research data supporting federal regulation.⁸⁴ Third and finally, scientists must share data in order to inform the scientific community of their own discoveries, accomplishments, and breakthroughs. This public notice serves to reward individual scientists through recognition and the esteem of their colleagues. Recognition has both intangible psychic benefits as well as tangible benefits, such as tenure advancement for junior investigators, increasingly prestigious appointments, endowed chairs, honorific positions for tenured investigators,⁸⁵ and perhaps most importantly in challenging economic times, scarce government grant funding.⁸⁶ According to the functionalist view, these three factors work in

83. See Dov Greenbaum, *Research Fraud: Methods for Dealing with an Issue that Negatively Impacts Society's View of Science*, 10 COLUM. SCI. & TECH. L. REV. 61, 75–77 (2009); Jocelyn Kaiser, *Data Integrity Report Sends Journals Back to the Drawing Board*, 325 SCI. 381 (2009) (citing several examples of scientific fraud, including a notoriously falsified South Korean stem cell paper in 2005, that prompted the commissioning of the NAS study and report cited *supra* note 82). The importance of making research data available for validation purposes was recently illustrated by the case of Dr. Marc Hauser at Harvard, whose conclusions regarding the behavior of tamarin monkeys were refuted by experts once they reviewed Hauser's original videotapes. Nicholas Wade, *In Harvard Lab Inquiry, a Raid and 3-Year Wait*, N.Y. TIMES, Aug. 13, 2010, at A12.

84. Omnibus Appropriations Act of 1999, Pub. L. 105-277, 112 Stat. 2681 (Data Access Act of 1999); Consolidated Appropriations Act of 2001, Pub. L. 106-554, § 515, 114 Stat. 2763, 2763A-154 (codified at 44 U.S.C. § 3516) (Data Quality Act of 2001). Despite the seemingly innocuous intent of this legislation, it has been criticized sharply for its potential to assist industry in challenging science-based governmental regulation. See, e.g., CHRIS MOONEY, *THE REPUBLICAN WAR ON SCIENCE* 102–20 (2005); Rick Weiss, *Data Quality Law Is Nemesis of Regulation*, WASH. POST, Aug. 16, 2004, at A1. See generally NAS, RESEARCH DATA, *supra* note 82, at 69.

85. Merton, *supra* note 5, at 274–75; see also NAS, RESEARCH DATA, *supra* note 82, at 55 (“Researchers receive intellectual credit for their work and recognition from their peers . . . when they publish their results and share the data on which those results are based.”); David, *supra* note 82, at 22.

86. In 2005, the National Institutes of Health (“NIH”), a major source of federal research funding, required that applicants have “demonstrated an ongoing record of accomplishments that have advanced their field(s).” Press Release, Nat’l Insts. of Health, *Enhancing Peer Review: The NIH Announces Enhanced Review Criteria for Evaluation of Research Applications Received for Potential FY2010 Funding*, Notice Number: NOT-OD-09-025 (Dec. 2, 2008), available at <http://grants.nih.gov/grants/guide/notice-files/not-od-09-025.html>. Commenting on the alterations to the NIH grant review process, a chairperson of an NIH grant review section noted that “applicants with robust publication histories, [and] proven track records of scientific accomplishment . . . may have the edge over their younger, less experienced counterparts.” Bob Grant, *New NIH Forms Raise Concerns*, THE SCIENTIST.COM (Dec. 8, 2009, 3:49 PM), <http://www.the-scientist.com/blog/display/56209/>. It should be noted, however, that the NIH does offer some types of grants specifically for less experienced, and thus less published, investigators. Nat’l Insts. of

concert, rewarding the individual scientist for sharing his or her data and yielding an overall benefit to the scientific community and society at large. As observed by the National Research Council, “the act of publishing [scientific findings] is a *quid pro quo* in which authors receive credit and acknowledgement in exchange for disclosure of their scientific findings.”⁸⁷

Recent commentators have questioned whether Merton’s norms accurately reflect the manner in which science is practiced today, arguing that competition and self-interest may motivate scientists far more than the larger social factors cited by Merton.⁸⁸ Patrick Taylor, who argues for broad data sharing in the biosciences on ethical and pragmatic grounds, worries that “research sharing is an aspiration vulnerable to compromise when it interferes with private incentives, whether researchers’ academic self-interest or potent market forces.”⁸⁹ Others have sought to explain scientific data sharing in economic terms by reference to the non-monetary incentives valued and earned by scientists (as individual economic actors) engaged in academic research.⁹⁰ Madison, Frischmann, and Strandburg, building on Ostrom’s institutional-choice foundation and functionalist intellectual property theory, posit that information commons “arise as solutions to collective action, coordination or transaction cost problems.”⁹¹

Health, *New and Early Stage Investigator Policies*, U.S. DEP’T OF HEALTH & HUMAN SERVS., http://grants.nih.gov/grants/new_investigators/ (last visited Sept. 17, 2010).

87. NAT’L RESEARCH COUNCIL, SHARING PUBLICATION-RELATED DATA AND MATERIALS: RESPONSIBILITIES OF AUTHORSHIP IN THE LIFE SCIENCES 34 (2003) [hereinafter NRC, SHARING PUBLICATION-RELATED DATA].

88. See generally HARRY COLLINS & TREVOR PINCH, THE GOLEM: WHAT EVERYONE SHOULD KNOW ABOUT SCIENCE (1993) (describing self-interested competition among scientists); IAN I. MITROFF, THE SUBJECTIVE SIDE OF SCIENCE (1974); Ian I. Mitroff, *Norms and Counter-norms in a Select Group of the Apollo Moon Scientists: A Case Study in the Ambivalence of Scientists*, 39 AM. SOC. REV. 579 (1974); Opderbeck, *supra* note 49, at 189–90 (summarizing various criticisms of Merton’s conceptualization of scientific norms). To his credit, Merton was not unaware of the intense competition and occasional skulduggery that characterize the scientific enterprise. In fact, he wrote extensively about these competitive factors, analyzing scientific rivalry and priority disputes since the great conflict between Newton and Leibnitz. See Merton, *supra* note 5, at 287; Robert K. Merton, *Behavior Patterns of Scientists (1968)*, in THE SOCIOLOGY OF SCIENCE, *supra* note 5, at 325, 336–37. The fact that individual scientists may act for reasons of self-interest, or that certain scientists may violate accepted norms, does not itself invalidate the observation of broader community norms.

89. Patrick I. Taylor, *Research Sharing, Ethics and Public Benefit*, 25 NATURE BIOTECHNOLOGY 398, 399 (2003).

90. See, e.g., Benkler, *supra* note 47, at 433–35; Peter Lee, *Toward a Distributive Commons in Patent Law*, 2009 WIS. L. REV. 917, 942.

91. Madison, Frischmann & Strandburg, *supra* note 23, at 691–92.

Another factor that has influenced the sharing of scientific data and the creation of science commons in recent decades is governmental policy. The federal government is by far the largest funder of basic scientific research in the United States.⁹² As such, federal policy plays a significant role in determining the manner in which scientific research is conducted. In general, current federal policy promotes the sharing and open accessibility of scientific data through a variety of contractual and grant-based mechanisms.⁹³ Federal policy thus reinforces and institutionalizes existing norms of scientific data sharing.⁹⁴

2. *Modes of Data Sharing*

The most typical and widespread means of disseminating results in the sciences is, and has been for centuries, publication in peer-reviewed scientific journals.⁹⁵ The quantity of a scientist's publications is one of the most important factors used in assessing the quality of his or her research, advancing his or her career, and determining his or her stature within the scientific community.⁹⁶ Thus, the scientist has a significant personal incentive

92. In 2008–2010, the annual federal budget for basic scientific research was approximately \$30 billion and represented approximately 60% of total basic research spending. JOHN F. SARGENT, JR., CONG. RESEARCH SERV., R40710, FEDERAL RESEARCH AND DEVELOPMENT FUNDING: FY2010, at 10 (2010), available at <http://www.fas.org/sgp/crs/misc/R40710.pdf>. Total federal spending on R&D, including basic and applied research, development and facilities, exceeded \$140 billion per year in 2008–2010. NAT'L SCI. FOUND., FEDERAL R&D FUNDING BY BUDGET FUNCTION: FISCAL YEARS 2008–2010 (2010), available at <http://www.nsf.gov/statistics/nsf10317/pdf/nsf10317.pdf>.

93. See Lee, *supra* note 90, at 941–42; Reichman & Uhler, *supra* note 61, at 331–32 (citing National Science Foundation data sharing requirements). Federal policy relating to genomic research is discussed in greater detail *infra* Section IV.A.

94. See Reichman & Uhler, *supra* note 61, at 332. Peter Lee also envisions governmental science policy as a means to achieve the ends of distributive justice, particularly in providing disadvantaged populations with access to health-promoting technologies. Lee, *supra* note 90.

95. See ZIMAN, *supra* note 4, at 39 (arguing that the peer-reviewed publication process is “at the very core of academic science” and “inseparable from its other functions”); Merton, *supra* note 88, at 337 (“From its very beginning, the journal of science introduced the institutional device of quick publication to motivate men of science to replace the value set upon secrecy with the value placed upon the open disclosure of the knowledge they had created.”).

96. See, e.g., Robert K. Merton, *Priorities in Scientific Discovery (1957)*, in *THE SOCIOLOGY OF SCIENCE*, *supra* note 5, at 286, 316 (noting the “tendency, in many academic institutions, to transform the sheer number of publications into a ritualized measure of scientific or scholarly accomplishment”); Editorial, *When Blogs Make Sense*, 466 NATURE 8 (2010) (“[P]eer-reviewed publications continue to be the primary markers of academic achievement.”).

to publish his or her results as quickly as possible⁹⁷ in as “high-impact” a journal as possible.⁹⁸ While in some disciplines researchers have begun to share scientific articles prior to publication through trusted “preprint” systems such as the arXiv.org server, publication in peer-reviewed journals remains an important and necessary step for recognition and validation.⁹⁹

The results and data reported in scientific journal articles, however, must be distinguished from the much larger quantity of experimental and observational data generated in the course of research and upon which published results are based.¹⁰⁰ A journal article typically includes a brief presentation of significant experimental findings, often made in summary or tabular fashion, together with the scientist’s analysis and conclusions based upon those findings.¹⁰¹ While the published data is usually essential to

97. A scientist’s incentive to publish quickly arises not only from a desire to advance his or her career, but also from a genuine fear that, if work is not published as soon as possible, competing groups may develop the same or similar results and “scoop” one’s work. See, e.g., NRC, SHARING PUBLICATION-RELATED DATA, *supra* note 87, at 28; Elizabeth Pennisi, *Genomics Researchers Upset by Rivals’ Publicity*, 329 SCI. 1585 (2010) (citing incidents of “scooping” in publishing sequencing results for various non-human genomes); Robert K. Merton, *Making It Scientifically*, N.Y. TIMES BOOK REV., Feb. 25, 1968, at 1 (reviewing James Watson’s *The Double Helix* and noting that “multiple independent discoveries are one of [the] occupational hazards” of science); *supra* note 3 and accompanying discussion.

98. See RESEARCH INFO. NETWORK, TO SHARE OR NOT TO SHARE: PUBLICATION AND QUALITY ASSURANCE OF RESEARCH DATA OUTPUTS 25 (2008), available at <http://www.rin.ac.uk/data-publication/> (noting that the assessment of researchers is “perceived to value above all else the publication of papers in high-impact journals”); ZIMAN, *supra* note 4, at 180 (noting that, in terms of scientific success, “[o]ne paper with a hundred favourable citations is worth infinitely more than a hundred papers with one citation each”).

99. See Editorial, *supra* note 96 (noting the popularity of preprint sharing of articles among astronomers, and its unpopularity among biologists); Donald Siegel & Philippe Baveye, Commentary, *Battling the Paper Glut*, 329 SCI. 1466 (2010) (criticizing the current importance placed on quantity of publications).

100. See Nelson, *supra* note 4, at 299.

101. See generally Rebecca S. Eisenberg, *Patents and Data-Sharing in Public Science*, 15 INDUS. & CORP. CHANGE 1013, 1024 (2006). By way of example, one recently published study identifies the *fgf4* gene as a factor leading to short-leggedness in dogs such as the Welsh corgi and the dachshund. Heidi G. Parker, *An Expressed Fgf4 Retrogene Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs*, 325 SCI. 995 (2009). The association of *fgf4* with the physical or “phenotypic” trait of short-leggedness is an experimental *result*. A vast quantity of *data* had to be collected and generated in order to arrive at this result, including raw genetic sequence reads for numerous dogs across different breeds, associated phenotypic data for each of the subjects, and a complex of statistical analyses, associations and computations.

It is also worth noting the distinction drawn by behavioral scientists among the terms *data*, *information*, and *knowledge*. Under the hierarchy developed by Fritz Machlup in 1983, “data” are individual facts, measurements, and assumptions (encompassing both “data” and “results,” as I have defined them), “information” is data that has been organized

support the scientist's analysis, the data reported in a journal article seldom represents the entirety of the "raw" data collected or observed by the scientist, and is typically only a small fraction of the full data set.¹⁰²

Traditionally, a scientist who wished to inspect another scientist's raw data, whether to validate the experimenter's results or to build upon those results, had only a few informal options for obtaining access to this data. Sometimes data would be presented at scientific conferences or symposia, though these venues typically provided the same (or less) data than would otherwise be presented in a published paper, albeit sooner. Many requests for experimental data were, and are still, made directly by one scientist to another by letter, telephone or, more recently, e-mail.¹⁰³ Such informal requests have typically been fulfilled, if at all, on the basis of friendship or professional courtesy, conditioned on the availability, work schedule, inclination, and discretion of the scientist in possession of the data.¹⁰⁴

Today, powerful electronic databases and high-speed networks that enable the dissemination of scientific data on an unprecedented scale supplement these informal and inefficient methods of data sharing. In the

and contextualized, and "knowledge" is information that has been processed and assimilated by a human. See DONALD O. CASE, LOOKING FOR INFORMATION: A SURVEY OF RESEARCH ON INFORMATION SEEKING, NEEDS, AND BEHAVIOR 64 (2d ed. 2007); Charlotte Hess & Elinor Ostrom, *Introduction: An Overview of the Knowledge Commons*, in KNOWLEDGE AS A COMMONS, *supra* note 18, at 3, 8. Due to the distinction I draw between "data" and "results," and because the presentation of scientific data regularly includes some analysis and synthesis, I do not adhere strictly to the Machlup hierarchy in my terminology, but rather use the term "information" to connote both "data" and "results," whether as individual data points or in an aggregated, processed form.

102. For example, the full genomic sequence of an organism might require hundreds or thousands of pages to print, whereas most journal articles are in the range of ten or fewer printed pages.

103. See Reichman & Uhler, *supra* note 61, at 343–48 (discussing the characteristics of informal data sharing arrangements between scientists).

104. See *id.* at 405 ("[R]esearchers . . . 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."). There is a growing body of empirical evidence demonstrating, however, that these requests for data sharing among scientists are often ignored or refused. See, e.g., Blumenthal et al., *National Survey*, *supra* note 17 (reporting that 8.9% of academic life scientists has refused to share research results with other scientists within the past three years); Blumenthal et al., *Prevalences and Predictors*, *supra* note 17 (concluding, on the basis of similar data to that presented in the authors' 2002 paper, that "data withholding is common in biomedical science"); Eric G. Campbell et al., *Data Withholding in Academic Genetics: Evidence from a National Survey*, 287 J. AM. MED. ASS'N 473 (2002) (reporting that 47% of geneticists who requested information relating to published research were denied at least once in the preceding three years, and 10% of all post-publication data results were denied); Editorial, *supra* note 3.

United States, the projects giving rise to these scientific commons have historically been “big science” initiatives in fields such as high-energy physics, astronomy, and geoscience, often using government facilities such as large telescopes, spacecraft, and particle accelerators.¹⁰⁵ More recently, however, scientific commons increasingly comprise data from projects that are generated by academic or research institutions funded in whole or in part by government grants.¹⁰⁶ In a typical arrangement of this nature, a government agency will fund these research centers to procure equipment and generate data in a coordinated or collaborative manner, either in fulfillment of a broader governmental program or as part of a research proposal made by the requesting institution.¹⁰⁷ The resulting data is often deposited in a government-operated database such as GenBank¹⁰⁸ and is thus made accessible to other scientists. This aggregation and availability of data in “science commons” enables the efficient, rapid, and cost-effective sharing of new knowledge and enables study and analysis that otherwise might have been impossible.¹⁰⁹

3. Stakeholders and Policy Considerations

In designing the “rules in use” for a science commons, policy makers must consider the interests of many different stakeholder groups, which may be overlapping, divergent, and sometimes contradictory.¹¹⁰ The principal stakeholder constituencies relevant to science commons are: (1) scientists who generate and contribute data to the commons (“data generators”); (2) academic or corporate institutions that employ these scientists;¹¹¹ (3)

105. See Reichman & Uhler, *supra* note 61, at 322.

106. Paul N. Schofield, et al., *Sustaining the Data and Bioresource Commons*, 330 SCI. 592, 592 (2010).

107. *Id.*

108. GenBank is operated by the National Library of Medicine of the National Institutes of Health.

109. See Eisenberg, *supra* note 101, at 1020. The system is not, however, without its flaws. See Keith Baggerly, Commentary, *Disclose All Data in Publications*, 467 NATURE 401 (2010) (noting that publicly available data is often insufficient to enable the reproduction of experiments in the field of genetics).

110. Both Ostrom and Hess, and Madison, Frischmann, and Strandburg emphasize the importance of identifying the various constituencies connected with a cultural commons. Madison, Frischmann & Strandburg, *supra* note 23, at 690; Ostrom & Hess, *supra* note 18, at 48–50.

111. While research scientists are often employees of their universities or other research institutions, individual scientists often exhibit different motivations and goals than their employer organizations. See Jennifer Carter-Johnson, *Unveiling the Distinction Between the University and Its Academic Researchers: Lessons for Patent Infringement and University Technology Transfer*, 12 VAND. J. ENT. & TECH. L. 473 (2010).

scientists who access and use the common data, and who in many cases are themselves data generators (“data users”); (4) academic or corporate institutions that employ the data users; (5) distributors and other intermediary handlers of the common data such as database operators, scientific journals, or in some cases, the data generators themselves; (6) funders of the research generating the data, such as government agencies, private foundations, or industry;¹¹² and (7) members of the general public, in their dual capacities as taxpayers who ultimately fund scientific research and as consumers of the products that eventually result from that research (and who, in the case of patient advocacy groups, may be highly conversant in the relevant scientific literature).¹¹³

Like the other “cultural commons” described by Madison, Frischmann, and Strandburg, science commons are constructs based on sets of rules.¹¹⁴ And as in other private ordering structures, the rules governing science commons can supplant baseline legal regimes such as intellectual property

112. See Lee, *supra* note 90, at 950–73 (addressing the interests and policy concerns of government and private funders of scientific research).

113. See *id.* at 986–90 (addressing the interests and policy concerns of disease advocacy groups); Lea Shaver, *The Right to Science and Culture*, 2010 WIS. L. REV. 121, 154–74 (proposing that public access to scientific and cultural information should be viewed as a basic human right within the framework of the Universal Declaration of Human Rights). Anupam Chander and Madhavi Sunder take issue with the fact that the “consumers” of information within scientific commons—and the genome commons, in particular—are “a self-selecting few who have chosen to participate in the new technology revolution.” Chander & Sunder, *supra* note 61, at 1343 (quoting Rodrigo Martinez et al., *The Geography of the Genome*, WIRED, June 2003, at 160). I do not suggest that the “commons” of scientific information is particularly valuable, or even comprehensible, to the majority of the world’s citizens. In many cases, the public interest in the broad dissemination of scientific information is represented by government or advocacy groups that possess the requisite qualifications to understand and use this data. This is not to say, however, that the public at large, which stands to benefit or lose based on the outcome of policy decisions, does not constitute an important “stakeholder” in the debate concerning the science commons.

114. Broadly speaking, information commons can arise organically or by design. Organic commons arise when informal groups share information about topics of common interest, as occurs in online communities that grow up around topics as diverse as cooking, sports, politics, music, open source software, childrearing, health, and online gaming. The information in these informal commons is generally contributed and used freely without restriction, though informal norms and collective behavioral expectations undeniably shape usage and production practices. See Benkler, *supra* note 47, at 381–82 (explaining how formal and informal norms structure collaboration on open source software development). Due to numerous organizational factors, including the capital-intensive nature of many data-generating scientific projects and the pervasiveness of governmental funding of scientific research, however, commons of scientific information are typically more structured.

protection to reflect the policy objectives of particular stakeholder groups.¹¹⁵ From the outset, it should be clear that the different stakeholder groups described above will have differing perspectives and incentives to act, and while each has a stake in the sharing of scientific data, no single group is motivated by all three of the Mertonian sharing incentives. Rather, each group responding to its own particular incentives and requirements ultimately results in the social sharing structures observed both by Merton and later functionalist scholars.

Data-generating scientists, for example, typically have a strong interest in publishing their work and obtaining grant funding to continue their research programs. The institutions that employ them, while concerned with the reputation and renown of their researchers, may have additional interests in obtaining patents or securing corporate funding (concerns that are often, but not always, shared by scientists, and not necessarily to the same degree). These interests may conflict with the interests of data users, who, as a gross generalization, typically wish to access and use generated data as soon as it is available with as little cost, inconvenience, and encumbrance as possible. Government funders may also seek rapid release of data to further the progress of science in the public interest, a concern that is often mirrored by patient advocacy groups. The conflicts that arise among these stakeholder groups are inevitable and will be addressed in greater detail in the case studies presented in Part IV, *infra*.

B. EXOGENOUS INFLUENCES ON LATENCIES WITHIN A SCIENCE COMMONS

As noted above, science commons are constructs developed through the application of rules. But privately ordered internal rule sets are not the only factors that influence the size and growth of science commons. In many cases, exogenous factors play a critical, if not deterministic, role in establishing the dynamic characteristics of a commons. In this Section, I examine three distinct exogenous factors that exert a substantial effect on knowledge and rights latency within a science commons¹¹⁶: (1) delays in the

115. See Peter Lee, *Contracting to Preserve Open Science: Consideration-Based Regulation in Patent Law*, 58 EMORY L.J. 889, 917 (2009) (discussing the increased control “shrinkwrap” licenses allow over content as compared to copyright and patent law).

116. There are, of course, numerous other factors that influence the design of information commons, particularly in the biomedical realm, including important considerations surrounding the protection of individual identity and health information and the proper means of obtaining informed consent. See, e.g., JOINT DATA STANDARDS STUDY, *supra* note 8, at 45–46; Kaye, *supra* note 36, at 333–35. While these considerations are critical

publication of data; (2) copyright and other legal constraints on the dissemination of data; and (3) patent encumbrances on the use of data.¹¹⁷

1. *Publication Delay*

The most common and regularized method of sharing scientific results is through publication in scientific journals.¹¹⁸ Typically, journals require that scientists provide sufficient detail regarding their methods and materials to permit a reader to reproduce and validate the experimental conclusions advanced in the article.¹¹⁹ In the past decade, however, an increasing number of scientific journals have required that authors make the underlying data supporting their published claims available to readers.¹²⁰ For example, in the case of genomic sequence data, journals often require deposit of the data into a public database at the time of publication.¹²¹ However, only a fraction of scientific journals have such requirements,¹²² and the requirements that do exist are often unclear about when, how and which data must be made available.¹²³ Thus, for most scientific papers, supporting data is published or

to the design and operation of information commons in the biomedical arena, they are relatively time-independent (i.e., subject privacy should be protected whenever data is released), and thus do not factor heavily into the latency analysis presented in this Article.

117. Interestingly, Madison, Frischmann, and Strandburg do not view science commons as particularly susceptible to encumbrance by intellectual property rights, suggesting that “commons of scientific research results and tools in the basic sciences” may be most appropriately characterized as “nonpropertized” realms that exhibit the characteristics of the “natural intellectual environment.” Madison, Frischmann & Strandburg, *supra* note 23, at 686. In this Part, I argue that intellectual property rights exert an extreme influence on science commons.

118. *See supra* Section III.A.2.

119. NRC, SHARING PUBLICATION-RELATED DATA, *supra* note 87, at 32.

120. *Id.* at 32–33; *see also* GUIDE TO PUBLICATION POLICIES OF THE NATURE JOURNALS (Apr. 30, 2009), *available at* <http://www.nature.com/authors/gta.pdf> (“[A] condition of publication in a Nature journal is that authors are required to make materials, data and associated protocols promptly available to readers without preconditions.” (emphasis removed)).

121. *See* Stephen Hilgartner, *Potential Effects of a Diminishing Public Domain in Biomedical Research Data*, in NRC, PUBLIC DOMAIN, *supra* note 7, at 137.

122. *See* NRC, SHARING PUBLICATION-RELATED DATA, *supra* note 87, at 33 (according to a 2003 study, of the fifty-six most frequently cited life-science and medical journals, only 41% had policies expressly requiring the deposit of data supporting published articles).

123. A journal’s data sharing requirement may be based on whether the data is *central* to the published results, *integral* to the published results, or simply useful *background* information. *See* NRC, SHARING PUBLICATION-RELATED DATA, *supra* note 87, at 36 (describing the “uniform principle for sharing integral data and materials expeditiously (UPSIDE)”).

made publicly available, if at all, no earlier than the time of publication.¹²⁴ The result, in most cases, is that a significant period of time elapses between the time that experimental data is generated and the time that it becomes publicly available. This delay may occur during one or more of the myriad steps in publishing results. Investigators must analyze their results, gather additional data, refine their analysis, prepare a paper based on their findings, and submit the paper to journals. Journals must conduct peer review and the editorial process, and wait for investigators to make any revisions required by the journals, including, at times, conducting additional experiments. If a journal rejects the paper, investigators must revise and submit it to different journals. Finally, the journal must edit, format, and prepare accepted papers for publication. One recent study reports that the period from completion of scientific work until publication is typically between twelve and eighteen months.¹²⁵ Older studies have found comparable or longer publication delay periods in other fields of research;¹²⁶ the seventeen-year delay in the case of the *Ardi* researchers represents an extreme example.¹²⁷

In terms of the latency analysis described in Section II.B.3, publication delay lengthens knowledge latency, often by years. The effect of this delay is complex. Though publication delay may hinder the overall progress of science (i.e., retarding other scientists' ability to build upon an investigator's findings), it does not slow the work of the investigators who initially collected the data—one of the criticisms leveled against the *Ardi* researchers. They, unlike their peers who lack access to this data during the publication process, are free to continue to analyze, refine, and conduct follow-up studies based on the data. Thus, even in an environment in which an ethic of data sharing has been widely accepted, this publication-based delay provides the original investigator with a substantial head-start in the analysis and

124. See, e.g., Reichman & Uhler, *supra* note 61, at 335; Toronto Int'l Data Release Workshop Authors, *Prepublication Data Sharing*, 461 NATURE 168, 168 (2009) [hereinafter Toronto Authors].

125. Carlos B. Amat, *Editorial and Publication Delay of Papers Submitted to 14 Selected Food Research Journals. Influence of Online Posting*, 74 SCIENTOMETRICS 379 (2008).

126. See William D. Garvey & Belver C. Griffith, *Scientific Information Exchange in Psychology*, 146 SCI. 1655, 1656 (1964) (reporting that in the psychology field, the time between hypothesis and publication is between 30 and 36 months, and the time between reportable results and publication is between 18 and 21 months); Charles G. Roland & Richard A. Kirkpatrick, *Time Lapse Between Hypothesis and Publication in the Medical Sciences*, 292 J. AM. MED. ASS'N 1273, 1274 (1975) (finding delays of 20 and 24 months between the completion of research and publication, respectively, for medical laboratory research and clinical research studies). Anecdotally, the author has been informed that publication delays are typically even longer in the social sciences.

127. See *supra* notes 1–3 and accompanying text.

publication of results based on his or her own experimental data. This distinction is important as it illustrates the dichotomy between two critical stakeholders in the science commons: data generators and data users. The divergent interests of these two constituencies played a major role in the development of the genome commons, as described in Section IV.A, *infra*.

2. *Copyright and Other Restraints on the Dissemination of Data*

Intellectual property protection has figured prominently in the ongoing debate over access to scientific data. Intellectual property can have two principal effects on an information commons: it can prevent the entry of data into the commons (impeding the growth of the overall knowledge base and thereby increasing knowledge latency), and it can limit the ability of users to utilize data that is already in the commons (impeding the growth of the unencumbered knowledge base and thereby increasing rights latency). Broadly speaking, intellectual property rights such as copyrights and trade secrets typically, though not always, fall into the first category and patents typically fall into the second category.

The effect of trade secret protection is the most straightforward to analyze. Scientific work that is sponsored by industry is often subject to written confidentiality obligations or other trade secret restrictions that explicitly prevent scientists from sharing resulting data and materials with others and, in some cases, delaying or even prohibiting the publication of their results.¹²⁸ BP's recent attempt to suppress data collected from the *Deepwater Horizon* oil spill is a prominent example of this practice.¹²⁹ With such restrictions in place, data cannot, practically speaking, enter a commons.¹³⁰ But corporate restrictions are not the only sources of secrecy in

128. See NAS, RESEARCH DATA, *supra* note 82, at 67; Allen C. Nunnally, *Intellectual Property Perspectives in Pharmacogenomics*, 46 JURIMETRICS J. 249, 255–58 (2006); Rai, *supra* note 4, at 111 (“Commercial involvement in academic research has . . . undermined norms governing the sharing of research materials and tools.”); INST. OF MEDICINE, EXTENDING THE SPECTRUM OF PRECOMPETITIVE COLLABORATION IN ONCOLOGY RESEARCH: WORKSHOP SUMMARY 36 (Margie Patlak, et al., eds. 2010) [hereinafter INST. OF MED.] (“Competing companies often compel their employees to keep silent about their endeavors, and the sharing of information is often frowned on lest information be divulged that might compromise the company’s competitive advantage.”).

129. See Mascarelli, *supra* note 16, at 538 (noting both the public outcry at BP’s initial attempts to “gag” scientific findings surrounding the leaking oil rig and subsequent attempts by the company to relax these restrictions).

130. At least not by legitimate means. Confidential data may be leaked or otherwise disclosed in violation of applicable confidentiality obligations and thereby enter the commons. This route toward public dissemination, however, is sporadic, unpredictable, and should not typically be relied upon by designers of information commons in the sciences.

the scientific community. Academic researchers themselves often have strong incentives to keep scientific data confidential, at least until the time of publication, and these incentives are supported, if not mandated, by university policies and procedures.¹³¹

The analysis of copyright restrictions is more subtle. Copyright law, unlike the law of trade secrets, does not have as its goal the concealment of information. Rather, copyright law grants the owner of a copyrighted work certain exclusive rights to exploit that work, including, for example, the exclusive rights to display, distribute, and reproduce.¹³² In the context of information commons, copyright principles are typically raised when discussing limitations on access to scientific data that has already been published. That is, even though scientific facts and conclusions are not themselves copyrightable, the articles (including text, diagrams and illustrations) in which they are presented are subject to copyright and thus controlled by the publishers of the journals carrying those articles.¹³³

In some cases, even data that might otherwise be in the public domain (such as mapping and geographic data developed under contract to the U.S. Federal Government) may be stored in proprietary databases that are accessible only to paid subscribers.¹³⁴ In several areas the “privatization” of governmental data is proceeding at a rapid pace due to (not entirely unjustified) perceptions of inefficiency and poor quality of governmental databases.¹³⁵ This situation, however, has led to fears that increasing amounts of data will be “enclosed” from public view; as Reichman and Uhler write, “the private sector simply cannot duplicate the government’s public good functions and still make a profit.”¹³⁶ This sentiment echoes a general criticism that the proprietary nature of scientific journals and electronic databases in which scientific results are published, and the high cost of subscribing to those journals and databases, is severely limiting the dissemination of

131. See INST. OF MED., *supra* note 128, at 36–37.

132. 17 U.S.C. § 106 (2006).

133. In many instances, authors are required to assign copyright in their articles and other works to journal publishers, or at least to grant publishers the exclusive right to publish their work in most relevant channels. See NAS, RESEARCH DATA, *supra* note 82, at 67.

134. See *id.* at 65. While the United States does not offer specific intellectual property protection for databases, such protection is available in Europe and elsewhere. See Reichman & Uhler, *supra* note 61, at 355.

135. See Reichman & Uhler, *supra* note 61, at 396.

136. *Id.* at 397.

knowledge.¹³⁷ The issue of scientific journals and the recent “open access” publishing movement in the sciences is discussed in greater detail in Section IV.B.

3. *Patents and Restraints on the Use of Data*
a) Effect of Patents on Scientific Data

Patents may be obtained in most countries to protect novel and inventive articles of manufacture, compositions of matter, and processes. Excluded from patentable subject matter are laws of nature and natural phenomena.¹³⁸ Thus, the information included within a science commons may be more or less susceptible of patent protection depending on its character. Pure scientific discoveries, such as observational data regarding the moons of Jupiter and conclusions regarding the giant planet’s mass and rotation drawn therefrom are unlikely to be suitable subject matter for patent protection. The same is not necessarily true in the biosciences, where the patentability of genetic, biological, and physiological associations has previously been affirmed by the U.S. Patent and Trademark Office and various courts (although the patentability of such subject matter remains the subject of substantial litigation and uncertainty).¹³⁹ In this respect, the significance of a

137. See, e.g., Susan R. Poulter, *Legal Pressures on the Public Domain: Licensing Practices*, in NRC, PUBLIC DOMAIN, *supra* note 7, at 101–03; Boyle, *supra* note 47, at 38–40; Reichman & Uhler, *supra* note 61, at 319–22. For a discussion of similar database protection issues in Europe and their potential effect on the dissemination of scientific knowledge, see NAT’L RESEARCH COUNCIL, BITS OF POWER — ISSUES IN GLOBAL ACCESS TO SCIENTIFIC DATA 150–53 (1997) [hereinafter NRC, BITS OF POWER]. See also Julie Cohen, *The Challenge of Digital Rights Management Technologies*, in NRC, PUBLIC DOMAIN, *supra* note 7, at 109 (discussing additional deleterious effects that DRM technologies may have on scientific research).

138. See, e.g., EUROPEAN PATENT OFFICE, EUROPEAN PATENT CONVENTION art. 52 (2000), available at <http://www.epo.org/patents/law/legal-texts/epc.html> (excluding from the meaning of invention, and thus patentable subject matter, “discoveries, scientific theories and mathematical methods”); JAPAN PATENT OFFICE, PART II: REQUIREMENTS FOR PATENTABILITY, CHAPTER 1: INDUSTRIALLY APPLICABLE INVENTIONS art. 1.1, available at http://www.jpo.go.jp/tetuzuki_e/t_tokkyo_e/Guidelines/2_1.pdf (noting that “a law of nature as such” is not considered to be a statutory invention). In the United States, patentable subject matter includes “any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof.” 35 U.S.C. § 101 (2006). This has been broadly interpreted by the United States Supreme Court to include “anything under the sun that is made by man.” *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980). However, the Court has also recognized exclusions to this broad interpretation, including “laws of nature, natural phenomena, and abstract ideas.” *Diamond v. Diehr*, 450 U.S. 175, 185 (1981).

139. See, e.g., *Bilski v. Kappos*, 130 S. Ct. 3218 (2010) (affirming the Federal Circuit’s finding that a simplistic business method claim of hedging risk in commodities was not patentable subject matter); *Prometheus Labs., Inc. v. Mayo Collaborative Servs.*, No. 09-490,

human genetic mutation is fundamentally different from, say, the eccentricity of Io's elliptical orbit: the scientific community's ability to utilize this information, even if it is publicly known, may differ greatly depending on the existence or absence of relevant patents. As such, the ability to use information resident within a commons of DNA cancer markers may be more constrained than the use of data within the commons of Jovian satellite information.¹⁴⁰ That is, when a valid patent claims an "invention" such as a means of diagnosing or treating a disease based on the presence of a specified genetic marker, the patent prevents any unauthorized person from "using" that information for such a purpose, even if the information is known to that person via the commons.

b) Patents and Latency

Unlike copyright and trade secret restrictions, which, in the contexts described above, tend to limit access to scientific data, patents are required to teach the "best mode" of enabling the inventions that they claim and thus, by their nature, disclose knowledge to the public. The U.S. Supreme Court has long held that the "ultimate goal of the patent system is to bring new . . . [ideas] and technologies into the public domain through disclosure."¹⁴¹ In the United States, most patent applications are published and become publicly available eighteen months after filing.¹⁴² The quid pro quo for public disclosure of the invention is the exclusive right to exploit the invention during the patent term (twenty years from filing in the United

2010 WL 2571881 (U.S. June 29, 2010) (remanding to the appellate court for reconsideration of whether simple medical diagnosis method claims meet the subject matter requirements of 35 U.S.C. § 101 in light of the Supreme Court's decision in *Bilski v. Kappos*, *supra*); Ass'n for Molecular Pathology v. U.S. Patent & Trademark Office, No. 09 Civ. 4515, 2010 WL 1233416 (S.D.N.Y. Mar. 29, 2010) (ruling that patent claims which claimed isolated human DNA sequences (genes) or methods of comparison of these genes to those of a human sample were ineligible for patent protection under 35 U.S.C. § 101 as they were products of nature and mental processes, respectfully).

140. Cf. INST. OF MED. & NAT'L RESEARCH COUNCIL, LARGE-SCALE BIOMEDICAL SCIENCE 27 (2003) [hereinafter LARGE-SCALE SCIENCE] (noting greater willingness among scientists working on "big-science" projects in astronomy and similar areas to share data due to the smaller commercial value of such discoveries as compared to that in the biosciences).

141. *Bonito Boats v. Thundercraft*, 489 U.S. 141, 151 (1989); see also LANDES & POSNER, *supra* note 54, at 294–99 (noting that the knowledge disclosed in a patent document may be used by competitors seeking to "invent around" the patented invention, but also observing that the competitive value of this knowledge decreases as the scope of patent protection increases); Eisenberg, *supra* note 101, at 1022–23 (explaining the argument that "disclosure of unprotected information is not an incidental byproduct of a process that aims to motivate something more worthwhile, but the whole point of the patent system").

142. 35 U.S.C. § 122 (2006).

States).¹⁴³ During this term, no person may make, use, sell, or import the patented invention without the permission of the patent owner, and even academic research activities have been found to constitute infringement.¹⁴⁴ Thus, in terms of latency analysis, patents create a “commons” of public knowledge in which knowledge latency may be expressed as the period of time between knowledge generation and publication of the patent application, and rights latency may be expressed as the period of time between publication of the application until expiration of the patent. Represented numerically, a “patent commons” would yield the following latency values:

$$\textit{knowledge latency (KL)} = \textit{preparation time} + 18 \textit{ months}$$

$$\textit{rights latency (RL)} = 240 \textit{ months} - 18 \textit{ months} = 222 \textit{ months}$$

These values are presented here principally for comparative purposes in the discussion that follows in Part V.

c) Patents and Anticommons

Significant debate exists concerning the effect that patents may have on scientific research and whether patents promote or chill innovation. While many commentators trace this controversy to the early 1980s and the enactment in the United States of the Bayh-Dole Act,¹⁴⁵ its roots extend at least to the 1850s and virulent opposition in Britain and continental Europe to the then-newly formalized patent systems.¹⁴⁶ Among the charges leveled

143. *Kewanee Oil v. Bicron*, 416 U.S. 470, 481 (1974). In this case, the Court states: [S]uch additions to the general store of knowledge are of such importance to the public weal that the Federal Government is willing to pay the high price of 17 years of exclusive use for its disclosure, which disclosure, it is assumed, will stimulate ideas and the eventual development of further significant advances in the art.

Id. (referring to the then-current seventeen-year patent term); *see also* *Sinclair & Carroll Co. v. Interchemical Corp.*, 325 U.S. 327, 330–31 (1945); *Griffith Rubber Mills v. Hoffar*, 313 F.2d 1 (9th Cir. 1963). *But see* Alan Devlin, *The Misunderstood Function of Disclosure in Patent Law*, 23 HARV. J.L. TECH. 401, 403–04 (arguing that patent disclosures seldom provide technically useful information to the public); Note, *The Disclosure Function of the Patent System (or Lack Thereof)*, 118 HARV. L. REV. 2007 (2005) (challenging the notion that in practice the patent system serves to disseminate knowledge).

144. *See* *Madey v. Duke Univ.*, 307 F.3d 1351 (Fed. Cir. 2002).

145. Bayh-Dole Act of 1980, Pub. L. No. 96-517, 94 Stat. 3015 (codified at 35 U.S.C. §§ 200–212). This controversial piece of legislation rationalized the previously chaotic rules governing federally sponsored inventions and allowed researchers to obtain patents on inventions arising from government-funded research.

146. *See* JOHNS, *supra* note 79, at 247–48, 267–75 (describing the mid-Victorian patent abolition movement).

against the patent system by Victorian critics was the deleterious effect that such grants of exclusive rights might have on the “commons” of scientific and technical knowledge.¹⁴⁷ Similar fears were expressed in the United States in the 1930s following the Federal Communications Commission’s investigation of AT&T and its massive patent holdings.¹⁴⁸

Anxiety over the effect of patents on the science commons continues today, the debate often being reduced to the oversimplified tradeoff between “access” and “incentive.”¹⁴⁹ In their influential 1998 paper, Michael Heller and Rebecca Eisenberg envision a scenario in which a proliferation of ownership claims within fields of study make it increasingly difficult to conduct research in those fields either because rights owners are unwilling to license patents on acceptable terms or because the sheer number of rights holders make it impractical or cost-prohibitive for researchers to procure the necessary rights.¹⁵⁰ They describe this effect as a “tragedy of the anticommons”¹⁵¹ and argue that such a state of affairs could effectively “privatize” research results that otherwise would have been freely available and thereby reduce the quantity of socially useful biomedical research.¹⁵² Other commentators have echoed Heller’s and Eisenberg’s concerns.¹⁵³ Arti Rai, in particular, argues that the increased availability of patents on inventions developed at academic laboratories (due in part to the enactment of the Bayh-Dole Act) has eroded the traditional “open science” norms of academic research and, consequently, has negatively affected the advancement of science.¹⁵⁴

147. *Id.* at 271.

148. *Id.* at 402–09 (calling the AT&T investigation “the era’s principal venue for debating the consequences of patents in general for society, science, and industry”).

149. See LANDES & POSNER, *supra* note 54, at 11.

150. Heller & Eisenberg, *supra* note 32. Heller and Eisenberg do not address the public-oriented approach of the Bermuda Principles described in Section IV.A, but focus instead on increasing patent-generating activity in other sectors of the biomedical research arena.

151. Echoing Hardin’s “tragedy of the commons.” See *supra* Section II.A.1.

152. Heller & Eisenberg, *supra* note 32, at 698.

153. See, e.g., Lori Andrews et al., *When Patents Threaten Science*, 314 SCI. 1395 (2006); Benkler, *supra* note 4, at 1110; Hilgartner, *supra* note 121, at 137 (“Excessive concern with the protection of intellectual property can erect barriers to establishing new public domains.”); Lee, *supra* note 115, at 903–05 (arguing that patents on basic biomedical research information have the potential to hinder research).

154. Rai, *supra* note 4, at 109 (“[U]niversities and individual researchers soon began to respond to the financial incentives of Bayh-Dole by rejecting communalism and increasing efforts to seek patents.”). Rai’s comprehensive analysis of norms in the biosciences is based, in part, on Eisenberg’s article. Rebecca S. Eisenberg, *Proprietary Rights and the Norms of Science in Biotechnology Research*, 97 YALE L.J. 177 (1987); see also DANIEL S. GREENBERG, SCIENCE

The effects and very existence of the “anticommons,” however, have been challenged by scholars who question many of the assumptions underlying Eisenberg’s and Rai’s construction of the “science commons” and counter that patents promote, rather than dampen innovation.¹⁵⁵ In the past few years, researchers have begun to apply empirical techniques to test the “tragedy of the anticommons.”¹⁵⁶ By and large these studies have been inconclusive, showing either no perceptible anticommons effect or only a modest effect.¹⁵⁷

FOR SALE: THE PERILS, REWARDS, AND DELUSIONS OF CAMPUS CAPITALISM 61 (2007); JENNIFER WASHBURN, UNIVERSITY, INC.: THE CORPORATE CORRUPTION OF HIGHER EDUCATION 73–75 (2005); John V. Frangioni, *The Impact of Greed on Academic Medicine and Patient Care*, 26 NATURE BIOTECHNOLOGY 503 (2008); Lee, *supra* note 115, at 941–42.

155. See, e.g., David E. Adelman, *A Fallacy of the Commons in Biotech Patent Policy*, 20 BERKELEY TECH. L.J. 985 (2005) (arguing that the public commons model incorrectly assumes that the commons for biomedical science is “finite and congested,” thus overstating the value of a commons approach); Shubha Ghosh, *How to Build a Commons: Is Intellectual Property Constrictive, Facilitative or Irrelevant?*, in KNOWLEDGE AS A COMMONS, *supra* note 18, at 209, 212–24 (arguing that the effect of intellectual property on commons development is context-dependent and, as such, may hinder, facilitate, or be irrelevant to commons building); F. Scott Kieff, *Facilitating Scientific Research: Intellectual Property Rights and the Norms of Science—A Response to Rai and Eisenberg*, 95 Nw. U. L. REV. 691 (2001) (rebutting the commons-based arguments posited by Rai and Eisenberg); Opderbeck, *supra* note 49, at 215–18 (arguing that private Coasian negotiations are likely to lead to optimal results in the market for biotechnology innovation).

156. See, e.g., John P. Walsh, Ashish Arora & Wesley M. Cohen, *Effects of Research Tool Patents and Licensing on Biomedical Innovation in Patents*, in PATENTS IN THE KNOWLEDGE-BASED ECONOMY 285 (Wesley M. Cohen & Stephen A. Merrill eds., 2003) (finding, based on interviews and archival data, little evidence that academic research has been impeded by concerns about patents on research tools); John P. Walsh, Charlene Cho & Wesley M. Cohen, *View from the Bench: Patents and Material Transfers*, 309 SCI. 2002 (2005) (finding that patents generally are not used to deny access to knowledge inputs, but that access to research materials may be more difficult); Karim R. Lakhani et al., *The Value of Openness in Scientific Problem Solving* (Harvard Bus. Sch., Working Paper No. 07-050, 2006), available at <http://www.hbs.edu/research/pdf/07-050.pdf> (offering empirical evidence in support of the proposition that the solution of scientific problems is facilitated by free and open information sharing); Fiona Murray & Scott Stern, *Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis* (NBER Working Paper Series, Paper No. 11465, 2005), available at <http://www.nber.org/papers/w11465/> (finding some empirical support for a modest anti-commons effect).

157. See, e.g., David, *supra* note 82, at 29 (“There is not much empirical evidence as to how altering the legal conditions and terms of intellectual property rights translates into change in the overall strength of economic incentives for the producers.”); Eisenberg, *supra* note 101, at 1018 (“It is less clear that patents are dampening incentives to engage in further basic research in the non-profit sector.”); Ann E. Mills & Patti M. Tereskerz, *Changing Patent Strategies: What Will They Mean for the Industry*, 25 NATURE BIOTECHNOLOGY 867, 867 (2007) (citing “diminished concern in the biotech industry over the possible occurrence of an anticommons”); Michael S. Mireles, *An Examination of Patents, Licensing, Research Tools, and the*

But even if an anticommons has not arisen as a result of patenting in the biosciences, the prospect of obtaining patent protection covering the fruits of scientific discoveries has had an effect on the rapidity with which scientific data is disclosed to the public. To ensure that disclosure and publication do not defeat patenting rights, many universities require that researchers submit proposed articles that disclose potentially patentable inventions to a university technology office for review prior to submitting the article to a journal in order to give the university sufficient time to prepare and file a patent application if so desired.¹⁵⁸ These express policies, together with increasing pressure from universities and research institutions to commercialize their scientific discoveries, have led to additional delays in the release of data, both through publication and other channels. One study has found that approximately twenty percent of life sciences faculty delayed publication of research results for more than six months due to patent or other commercial issues.¹⁵⁹ Another study reports that scientists with an

Tragedy of the Anticommons in Biotechnology Innovation, 38 U. MICH. J.L. REFORM 141, 171–94 (2005) (finding previous studies inconclusive); Charles R. McManis & Suheol Noh, *The Impact of the Bayh-Dole Act on Genetic Research and Development: Evaluating the Arguments and Empirical Evidence to Date 48* (Aug. 13, 2006) (unpublished draft), available at <http://www.law.berkeley.edu/institutes/bclt/ipsc/papers2/mcmanis.doc> (“[V]ery little empirical evidence has been produced to date to support the argument that granting patents on the results of ‘upstream’ genetic research undermines the norms of the biological research community or retards biomedical innovation, technology transfer, or the development of downstream commercial products and processes.”). Indeed, some economics scholars even point out that the historical “enclosure” movement that fenced-off common pastures in England from the sixteenth to nineteenth centuries, and which is often cited by commons proponents as a negative development, see, for example, Boyle, *supra* note 47, significantly enhanced agricultural output. LANDES & POSNER, *supra* note 54, at 12.

158. Frangioni, *supra* note 154, at 504 (describing the rise of academic invention disclosure policies geared toward patent filings); see, e.g., *About WUSTL: Compliance & Policies: Intellectual Property Policy*, WASHINGTON U. IN ST. LOUIS sec. I.3.d (Feb. 15, 2008), <http://wustl.edu/policies/intelprop.html>. Section I.3.d states:

Nothing in this Policy shall limit or restrict the right of University faculty and students to publish results of their research, subject to reasonable delays to preserve patent or other intellectual property rights. Delays in publication required by the University . . . as a general rule, shall not exceed ninety days from initial disclosure of the intellectual property to the Research Office

Id.; see also UNIV. OF N.C. AT CHAPEL HILL, PATENT & INVENTION POLICY sec. III (2009), available at <http://www.unc.edu/campus/policies/patent%20policy%2000015747.pdf> (“[T]his Patent & Invention Policy does not limit the right to publish, except for short periods of time necessary to protect patent rights.”).

159. Blumenthal et al., *National Survey*, *supra* note 17, at 1224. This data has been supported by numerous subsequent studies, including one reporting that the disclosure of a gene associated with the condition hemochromatosis (an iron overload disease) was delayed

interest in patenting their results regularly withhold presentation of their data at scientific meetings.¹⁶⁰

In this Article, I do not seek to stake out a normative position regarding the effect, positive or negative, that patents and other intellectual property rights have on innovation and scientific advancement. I do, however, wish to draw attention to the very real and measurable effect that patents and other intellectual property have on science commons. That is, these forms of intellectual property protection, whether they promote or deter scientific research, undeniably reduce the rate at which the results of such research are contributed to the store of public knowledge and the speed at which such information becomes generally usable. Whether or not this delay yields a net benefit to society, policy makers designing information commons in the sciences will likely act in accordance with their own perceptions regarding this normative question. In the case studies that follow, I examine two well-known narratives of commons formation in the sciences with a particular view toward the timing-based adjustments and compromises that policy designers implemented in furtherance of their own institutional goals.

IV. TIMING AND DATA SHARING: TWO CASE STUDIES

A. THE GENOME COMMONS¹⁶¹

1. *Development of Genomics Data Release Policies*

a) The Human Genome Project (HGP)

The Human Genome Project, one of the most ambitious scientific undertakings in history, has been compared in scope and importance to the Manhattan Project and the manned space program.¹⁶² The project, which

by more than a year after the first patent application was filed by the sponsoring institution. Jon F. Merz et al., *Diagnostic Testing Fails the Test: The Pitfalls of Patents are Illustrated by the Case of Haemochromatosis*, 415 NATURE 577 (2002).

160. Jeremy M. Gruschcow, *Measuring Secrecy: A Cost of the Patent System Revealed*, 33 J. LEGAL STUD. 59, 73 (2004) (measuring the gap between patent filing and presentations at cancer research meetings).

161. A more detailed analysis of the policy considerations surrounding the development of the “genome commons” can be found in Contreras, *supra* note 35.

162. *See, e.g.*, FRANCIS S. COLLINS, *THE LANGUAGE OF LIFE* 2 (2010); ARTHUR M. LESK, *INTRODUCTION TO GENOMICS* 22 (2007); James D. Watson, *The Human Genome Project: Past, Present and Future*, 248 SCI. 44, 44 (1990).

spanned fifteen years and involved over a thousand scientists worldwide, has resulted in major advances in biochemistry, bioinformatics, and genetics.¹⁶³

From the outset, scientists expected that the HGP would generate large quantities of data regarding the genetic make-up of humans and other organisms. Thus, in 1988 the National Research Council was commissioned to study issues likely to be raised by the HGP. It recommended that all data generated by the project “be provided in an accessible form to the general research community worldwide.”¹⁶⁴ Other influential groups and policy makers soon took up the call for the commitment of all human sequence information to the public domain.¹⁶⁵ Perhaps most importantly, many leading scientists involved in the project urged that the output of the HGP be released to the public in the service of science.¹⁶⁶ They may have been influenced by pre-existing data sharing norms in the research communities in which they worked,¹⁶⁷ or held a genuine belief that genomic information, as the common biological heritage of the human species, was qualitatively *different* than other scientific data. As expressed by Dr. Ari Patrinos, the Department of Energy’s (DOE) Associate Director for Biological and Environmental Research, “the genome belongs to everybody.”¹⁶⁸

163. See generally COLLINS, *supra* note 36, at 674 (describing the “profound impact on scientific progress” achieved by the HGP); Int’l Hum. Genome Sequencing Consortium, *Initial Sequencing and Analysis of the Human Genome*, 409 NATURE 860, 911–13 (2001) [hereinafter Genome Sequencing Consortium].

164. NAT’L RESEARCH COUNCIL, MAPPING AND SEQUENCING THE HUMAN GENOME 8 (1988) [hereinafter NRC, HUMAN GENOME].

165. See NAT’L RESEARCH COUNCIL, REAPING THE BENEFITS OF GENOMIC AND PROTEOMIC RESEARCH 27 (2006) [hereinafter NRC, GENOMIC AND PROTEOMIC RESEARCH] (citing a 1988 NIH advisory council report stating that “the data must be in the public domain, and the redistribution of the data should remain free of royalties”).

166. James Watson, then director of the National Center for Human Genome Research, wrote in 1990 that “making the sequences widely available as rapidly as practical is the only way to ensure that their full value will be realized.” Watson, *supra* note 162, at 48; see also COLLINS, *supra* note 162, at 301; Boyle, *supra* note 47, at 37.

167. See Robert Cook-Deegan, *The Urge to Commercialize: Interactions Between Public and Private Research Development*, in NRC, PUBLIC DOMAIN, *supra* note 7, at 89 (“There were [] communities doing molecular biology at the same time on yeast and *Drosophila* that had ‘open-science’ norms. Those norms were the ones adopted as the models for the Human Genome Project.”); Genome Sequencing Consortium, *supra* note 163, at 864. The evolution of the open science culture among *c. elegans* researchers is described in some detail in NRC, GENOMIC AND PROTEOMIC RESEARCH, *supra* note 165, at 54–56.

168. Eliot Marshall, *Bermuda Rules: Community Spirit, With Teeth*, 291 SCI. 1192 (2001).

In 1992, the federal co-managers of the project, NIH¹⁶⁹ and DOE, approved joint guidelines for the sharing of data by investigators performing HGP-funded activities, acknowledging that the rapid sharing of HGP materials and data was “essential for progress toward the goals of the program,” and that such sharing would avoid unnecessary duplication of effort and expedite research in other areas.¹⁷⁰ Genomic sequence data generated by the HGP would be deposited in the NIH-managed GenBank database, and would be accessible to all scientists.¹⁷¹ However, the 1992 guidelines counterbalanced the desire for rapid public release of data against a perceived need to give researchers “some scientific advantage from the effort they have invested.”¹⁷² In recognition of this effort, the 1992 guidelines established a six-month maximum period from the time that HGP data is generated until the time that it must be made publicly available (i.e., knowledge latency of six months).¹⁷³ During this six-month latency period, HGP researchers could analyze their data and prepare publications, and only after the end of the six-month period were they required to release the data into public databases.¹⁷⁴

169. NIH formed the National Center for Human Genome Research (NCHGR) in 1989, under the direction of James Watson, to carry out its component of the HGP. In 1997, the Department of Health and Human Services elevated NCHGR to the status of a full “institute” within the NIH system, forming the National Human Genome Research Institute (NHGRI). *About the Institute: A History and Timeline*, NAT’L HUM. GENOME RES. INST., <http://www.genome.gov/10001763/> (last visited Sept. 17, 2010).

170. *DOE-NIH Guidelines Encouraging Sharing of Data and Resources*, HUM. GENOME NEWS, Jan. 4, 1993, at 4 [hereinafter *DOE-NIH Guidelines*].

171. See NRC, GENOMIC AND PROTEOMIC RESEARCH, *supra* note 165, at 4–5.

172. *DOE-NIH Guidelines*, *supra* note 170, at 4.

173. This six-month period was significantly shorter than release periods for other government-funded projects, which often allowed researchers to retain their data privately either until publication of results or for some specified “exclusivity period,” usually in the neighborhood of twelve to eighteen months. See, e.g., NAS, RESEARCH DATA, *supra* note 82, at 64 (noting that NASA and the European Southern Observatory Administration impose twelve-month proprietary periods and the U.S. National Optical Astronomy Observatory imposes an eighteen-month proprietary period on the release of data); NRC, SHARING PUBLICATION-RELATED DATA, *supra* note 87, at 75 (describing the one-year “hold allowance” on the deposition of crystallography data into the Protein Data Bank); NRC, BITS OF POWER, *supra* note 137, at 80–82 (describing data release policies of NASA and Global Change Research Program); Reichman & Uhler, *supra* note 61, at 335 (“In most cases, publication of research results marks the point at which data produced by government-funded investigators should become generally available.”).

174. *DOE-NIH Guidelines*, *supra* note 170.

b) The Bermuda Principles

By 1996, the HGP, together with its international collaborators, was ready to begin sequencing the human genome. In February, approximately fifty scientists and policy makers met in Hamilton, Bermuda to discuss the sharing and release of HGP-funded genomic data.¹⁷⁵ Among the issues debated was the speed at which data should be released to the public, and whether the six-month “holding period” approved in 1992 should continue.¹⁷⁶ Several arguments for eliminating the holding period were presented. From a pragmatic standpoint, scientists argued that gene sequencing centers working on the HGP required regularly updated data sets in order to avoid duplication of effort and to optimize coordination of the massive, multi-site project.¹⁷⁷ Waiting six months to obtain data was simply not practical if the project were to function effectively. But perhaps more importantly, the concept of rapid data release became endowed with an ideological character: the early release of data was viewed as necessary to accelerate the progress of science.¹⁷⁸ The result of the meeting was the

175. The meeting, sponsored by the Wellcome Trust, was designated the International Strategy Meeting on Human Genome Sequencing and held on February 25–28, 1996. Participants included representatives of NIH and DOE, the Wellcome Trust, UK Medical Research Council, the German Human Genome Programme, the European Commission, the Human Genome Organisation (HUGO) and the Human Genome Projects of France and Japan. See David Smith & Anthony Carrano, *International Large-Scale Sequencing Meeting*, HUM. GENOME NEWS, Apr.–June 1996, at 19.

176. See Marshall, *supra* note 168; Robert Mullan Cook-Deegan & Stephen J. McCormack, *A Brief Summary of Some Policies to Encourage Open Access to DNA Sequence Data*, 293 SCI. 217 (2001) (supplemental material), available at <http://www.sciencemag.org/cgi/content/full/293/5528/217/DC1>.

177. David R. Bentley, *Genomic Sequence Information Should Be Released Immediately and Freely in the Public Domain*, 274 SCI. 533, 533 (1996); see also Adam Bostanci, *Sequencing Human Genomes*, in FROM MOLECULAR GENETICS TO GENOMICS 174 (Jean-Paul Gaudillière & Hans-Jörg Rheinberger eds., 2004).

178. See, e.g., Bentley, *supra* note 177, at 534; Francis Collins et al., *A Vision for the Future of Genomics Research*, 422 NATURE 835, 846 (2003) (“Scientific progress and public benefit will be maximized by early, open and continuing access to large data sets.”); Cook-Deegan & McCormack, *supra* note 176 (“[W]ithout [the Bermuda Principles] the wait for information sufficient to meet patent criteria from high throughput sequencing programs would lead to long delays, and thus be a serious drag on science, undermining the publicly funded sequencing programs’ very purpose.”); Genome Sequencing Consortium, *supra* note 163, at 864 (“We believed that scientific progress would be most rapidly advanced by immediate and free availability of the human genome sequence. The explosion of scientific work based on the publicly available sequence data in both academia and industry has confirmed this judgment.”); Lee, *supra* note 115, at 905 (referring to “NIH’s aggressive intervention to enhance access to taxpayer-financed research tools”); Toronto Authors, *supra* note 124, at 169–70.

“Bermuda Principles,” a set of guidelines endorsed by each participant and intended to prevent any genome sequencing center from establishing a “privileged position in the exploitation and control of human sequence information.”¹⁷⁹ Most significantly, the Bermuda Principles required that genomic data be released to GenBank within twenty-four hours following assembly, the fastest data release requirement in history.¹⁸⁰

c) Bermuda, Data Release, and Patents

Not all observers viewed the adoption of the Bermuda Principles as an unmitigated success. Some critics questioned the effect that rapid release of data would have on scientists’ ability to publish the results of their work before others who had accessed it via public means.¹⁸¹ And, more pragmatically, some claimed that rapid data release was not even required in order to achieve the scientific goals of the HGP.¹⁸² But the most persistent skepticism regarding the Bermuda Principles came from those who believed that they subverted researchers’ ability to protect their discoveries with patents. In particular, the Bermuda Principles and their adoption by NIH in 1997 ensured that genomic data from the HGP and other large-scale sequencing projects would be made publicly available before data generators had an opportunity to file patent applications covering any “inventions” arising from that data, and in a manner that ensured its availability as prior art against third party patent filings at the earliest possible date.¹⁸³ As a result, the ability of both data generators and third parties to obtain patents on HGP

179. Bermuda Principles, *supra* note 35.

180. *Id.*

181. Deanna M. Church & LeDeana W. Hillier, *Back to Bermuda: How Is Science Best Served?*, 10 GENOME BIOLOGY 105, 105.1 (2009) (“[T]here was some concern that [the policy] would jeopardize the genome center’s ability to analyze and publish the data they had produced.”).

182. Bostanci, *supra* note 177, at 175 (arguing that the yeast genome project, a cooperative effort of seventy-nine different laboratories, operated successfully with a delayed data release policy).

183. In jurisdictions such as the European Union and Japan that have so-called “absolute novelty” requirements, an invention may not be patented if it has been publicly disclosed prior to the filing of a patent application. *See* JOHN GLADSTONE MILLS III ET AL., PATENT LAW FUNDAMENTALS sec. 2:30 (perm. ed., rev. vol. 2009). In such countries, a description of the invention in a scientific journal could preclude the inventor from obtaining patent protection for his or her invention. In the United States, a patent application may be filed with respect to an invention that has been disclosed in a printed publication, but only if the publication occurred less than one year before the filing of the patent application. 35 U.S.C. § 102(b) (2006). Thus, an inventor seeking patent protection for his or her invention must file a patent application prior to the disclosure of the invention in a publication (or, in the United States, no more than one year following publication).

data was severely curtailed, both on grounds of “novelty” (one cannot patent material that is “known . . . by others”)¹⁸⁴ and “nonobviousness” (one cannot patent material that is obvious in view of the prior art).¹⁸⁵ Rapid release of HGP data under the Bermuda Principles thus achieved a second policy goal of project leaders: limiting patent-related encumbrances on HGP data and the fruits thereof.¹⁸⁶ Critics of this approach charged that the NIH’s support of the Principles contravened the requirements of the Bayh-Dole Act, which expressly permits federally funded researchers to patent their discoveries.¹⁸⁷ However, scientists and NIH leaders dismissed these arguments regarding patents and Bayh-Dole, arguing instead that patenting of raw DNA sequences is inappropriate and should be impossible.¹⁸⁸

184. 35 U.S.C. § 102(a). This approach has also been used in the case of traditional knowledge released by the Indian government through a publicly accessible digital library for the purpose of constituting prior art “to prevent erroneous patent grants.” John Swinson & Laura Pearson, *Protection of Traditional Yoga-Related Knowledge in a Modern World*, BNA WORLD INTELLECTUAL PROP. RPT., Sept. 3, 2010.

185. 35 U.S.C. § 103 (2006). For example, the publication of a full genetic sequence would likely prevent the patenting of a gene included in that sequence even though the gene is disrupted by gaps (introns) in the genetic code that are excised during gene expression. Alexander K. Haas, *The Wellcome Trust’s Disclosure of Gene Sequence Data into the Public Domain & the Potential for Proprietary Rights in the Human Genome*, 16 BERKELEY TECH. L.J. 145, 158–59 (2001).

186. *See supra* note 166. As a corollary effect, the requirement of the Bermuda Principles that sequence data be deposited into GenBank and other public databases also had the effect of eliminating the copyright barriers to data sharing that are discussed *supra* notes 122–37 and accompanying text. Interestingly, Rebecca Eisenberg suggests that, in some cases, the early release of experimental data may actually encourage *more* patent filings by third parties who are thereby enabled to combine public data with proprietary improvements and patent the combination thereof. Eisenberg, *supra* note 101, at 1026.

187. *See* JAMES SHREEVE, *THE GENOME WAR* 46 (2004) (“Strictly speaking, the policy directly contradicted the Bayh-Dole Act.”); Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 LAW & CONTEMP. PROBS. 289, 308 (2003) (“Arguably, NIH has acted outside the scope of its statutory authority . . . at least with respect to patentable inventions.”).

188. Bentley, *supra* note 177, at 534; *see also* JAMES D. WATSON ET AL., *RECOMBINANT DNA* 295 (3d ed. 2005); Marshall, *supra* note 168. The NIH policy implementing the Bermuda Principles states that “raw human genomic DNA sequence, in the absence of additional demonstrated biological information, lacks demonstrated specific utility and therefore is an inappropriate material for patent filing.” *NHGRI Policy Regarding Intellectual Property of Human Genomic Sequence*, NAT’L HUM. GENOME RES. INST., Apr. 9, 1996, <http://www.genome.gov/10000926/> [hereinafter *NHGRI 1996 Policy*]. NIH’s position regarding the patenting of genetic material can be traced to its abortive attempt to seek patents on short genetic sequences known as expressed sequence tags (ESTs). In early 1991, NIH filed patent applications covering hundreds of ESTs and announced its intention to continue seeking such patents in the future. *See* Thomas Barry, *Revisiting Brenner: A Proposed Resolution to the Debate Over the Patentability of Expressed Sequence Tags Using the Concept of Utility*

d) Growing Tension Between Data Generators and Users

Despite the success and broad adoption of the Bermuda Principles, in the years immediately following the completion of the HGP a number of large-scale, publicly funded genomics projects began to adopt data release policies reflecting an increasing recognition of the inherent tension between data *generators*, the scientists at large-scale gene sequencing centers that produce the bulk of raw genomic data, and data *users*, those scientists who access and use genomic data to investigate specific disease-gene associations and conduct other analyses that rely on this data. In 2003, the Wellcome Trust sponsored a meeting in Ft. Lauderdale, Florida to revisit issues of pre-publication data release in the “post-genome” world.¹⁸⁹ The most significant result of the Ft. Lauderdale meeting was an agreement that a broad category of scientific projects, termed “community resource projects” (CRPs), should be subject to the rapid, pre-publication data-release rules established in Bermuda.¹⁹⁰ NIH’s 2003 genomic data release policy, issued just a few weeks after the Ft. Lauderdale meeting, reiterated the agency’s commitment to the Bermuda Principles and recognized the need for data generators to publish analyses of their data.¹⁹¹ Yet despite this acknowledgement, the agency

Control, 35 AIPLA Q.J. 1, 11 (2007); Rebecca Eisenberg, *Intellectual Property at the Public-Private Divide: The Case of Large-Scale cDNA Sequencing*, 3 U. CHI. L. SCH. ROUNDTABLE 557, 558–59 (1996). This announcement sparked what Robert Cook-Deegan refers to as “an international firestorm” and led to threats by other HGP collaborators to refuse to share their own data with NIH. ROBERT COOK-DEEGAN, *THE GENE WARS: SCIENCE, POLITICS, AND THE HUMAN GENOME* 330–31 (1994). In addition to negative reactions from abroad, the NIH’s and DOE’s own advisory committees were “unanimous in deploring the decision to seek such patents.” *Id.* at 317. Even James Watson decried the EST patenting plan as “sheer lunacy.” SHREEVE, *supra* note 187, at 84–85. The EST debate marked a turning point in NIH’s attitude toward patents on genetic material. Despite its initial enthusiasm for such patents, in 1994 NIH elected not to appeal the Patent and Trademark Office’s rejection of its EST patent applications. *See* LARGE-SCALE SCIENCE, *supra* note 140, at 36–37. After that, the PTO adopted a consistently lukewarm, if not outright averse, attitude toward the patenting of genetic sequences. *See* NRC, *GENOMIC AND PROTEOMIC RESEARCH*, *supra* note 165, at 52–53.

189. THE WELLCOME TRUST, *SHARING DATA FROM LARGE-SCALE BIOLOGICAL RESEARCH PROJECTS: A SYSTEM OF TRIPARTITE RESPONSIBILITY* (2003), available at <http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf> [hereinafter Ft. Lauderdale Principles].

190. As defined in the Ft. Lauderdale Principles, a “community resource project,” like the original HGP, is “a research project specifically devised and implemented to create a set of data, reagents or other material whose primary utility will be as a resource for the broad scientific community.” *Id.* at 2–3.

191. *Id.* at 4 (“[T]he sequence producers have a legitimate interest in publishing peer-reviewed reports describing and analyzing the sequence they have produced.”).

declined to impose any explicit limitation on users of the released sequence data. Instead, the policy reaffirmed NIH's position that DNA sequence data "should be available for all to use without restriction" and urged data users to act within "standard scientific norms" and to acknowledge data generators in published analyses based on their data.¹⁹² Data generators were left with the unsatisfying conclusion that potential loss of their publication priority was "a *necessary risk* set against the considerable benefits of immediate data release."¹⁹³

e) GWAS and Embargo Policies

Between 2003 and 2006, the technologies available for genomic research continued to improve in quality and decrease in cost, resulting in an expansion of the scope of genomic research and the advent of so-called genome-wide association (GWA) studies.¹⁹⁴ These studies differ from pure sequencing projects in that their goal is not the generation of large data sets (such as the genomic sequence of a particular organism), but the discovery of disease markers or associations hidden within the genome. They thus have greater potential clinical utility and lie further along the path to ultimate commercialization as therapeutics or diagnostics. Moreover, as the types of data involved in large-scale genomics projects expanded, the community of researchers participating in these projects became more diverse. Today, many scientists with backgrounds outside of genomics, including medical researchers, medical geneticists, clinicians, and epidemiologists, actively lead and participate in GWA projects. Yet these scientists do not necessarily share the norms of rapid pre-publication data release embraced by the genomics community since the early days of the HGP.¹⁹⁵ In many cases, particularly when patient data is involved, these scientists are accustomed to an

192. *Id.* For a general critique of the NIH's "hortatory" approach to this issue, see Rai & Eisenberg, *supra* note 187, at 293–94.

193. Ft. Lauderdale Principles, *supra* note 189, at 4 (emphasis added). In addition, the Ft. Lauderdale Principles suggest that data generators publish "project descriptions" (sometimes referred to as "marker papers" or "statements of intent") in order to establish a published priority claim to the work they are conducting. *See id.* at 3; Toronto Authors, *supra* note 124, at 169. Based on informal discussions with genomic scientists, it appears that few of these project descriptions have been published, though efforts are emerging to increase this practice. *See Editorial: Cite Site*, 42 NATURE GENETICS 729 (2010) (describing the *Nature Preprints* archive that is intended to offer a place for data generators to list citable project descriptions (marker papers) and receive citation credit).

194. *See* Monya Baker, *Genome Studies: Genetics by Numbers*, 451 NATURE 516 (2008).

195. *See, e.g.*, Cook-Deegan, *supra* note 167, at 90 (noting "thousands of labs" outside the traditional genomics disciplines that perform gene sequencing, but few of which observe the Bermuda Principles).

environment in which data is tightly guarded and released only after publication of results, and then only in a limited, controlled manner.

Accordingly, when the federally funded Genetic Association Information Network (GAIN) was established in 2006 to conduct GWA studies of six common diseases, the program's data release policies reflected a compromise among data generators and data users.¹⁹⁶ Data generators agreed to "immediate" release of data generated by the project, but for the first time a temporal restriction was also placed on *users* of the data.¹⁹⁷ That is, in order to secure a period of exclusive use and publication priority for data generators, data users were prohibited from submitting abstracts and publications and making presentations based on GAIN data for a specified "embargo" period¹⁹⁸ generally fixed at nine months.¹⁹⁹ This embargo period is an example of rights latency.

Shortly thereafter, NIH adopted a similar embargo-based approach in its institute-wide policy regarding the generation, protection and sharing of data generated by federally funded GWA studies (NIH GWAS Policy).²⁰⁰ The NIH GWAS Policy states that users of GWA data should refrain from submitting their analyses for publication, or otherwise presenting them publicly, during an "exclusivity" period of up to twelve months from the date that the data set is first made available.²⁰¹ While the agency expresses a "hope" that "genotype-phenotype associations identified through NIH-supported and NIH-maintained GWA data sets and their obvious implications will remain available to all investigators, unencumbered by

196. See generally The GAIN Collaborative Research Grp., *New Models of Collaboration in Genome-Wide Association Studies: The Genetic Association Information Network*, 39 NATURE GENETICS 1045 (2007).

197. *Id.* at 1048 box 1.

198. GENETIC ASS'N INFO. NETWORK (GAIN), DATA USE CERTIFICATION AGREEMENT 4–5 (Dec. 3, 2008) [hereinafter GAIN DATA USE AGREEMENT], available at http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?view_pdf&stacc=phs000016.v1.p1 (Terms of Access, no. 8).

199. *Id.*; The GAIN Collaborative Research Grp., *supra* note 196, at 1049.

200. See Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS), 72 Fed. Reg. 49,290 (Aug. 28, 2007) [hereinafter NIH GWAS Policy].

201. *Id.* at 49,296. This exclusivity period was originally proposed to be nine months, but was subsequently lengthened following NIH's receipt of public comments from data generators and others. See Request for Information (RFI): Proposed Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS), 71 Fed. Reg. 51,629 (Aug. 30, 2006).

intellectual property claims,”²⁰² the agency stops short of prohibiting the patenting of resulting discoveries.

Embargo mechanisms have also been implemented in subsequent genomic and related data release policies.²⁰³ At a 2009 summit in Toronto, more than one hundred scientists, scholars, and funding agency representatives met to assess the current state of rapid pre-publication data release and the applicability of the Bermuda Principles to projects well beyond the generation of genomic sequence data.²⁰⁴ The participants reaffirmed a general commitment to rapid pre-publication data release, but at the same time acknowledged the acceptability of a “protected period” (rights latency) of up to one year during which data users may be restricted from publishing analyses of released data sets.²⁰⁵

f) Private Sector Policies

A noteworthy parallel to the government-sponsored projects discussed above is that of private-sector initiatives in the genome sciences. Pharmaceutical giant Merck organized the first of these in 1994, when it established a large public database of expressed sequence tags (ESTs). The stated purpose of the so-called Merck Gene Index was to increase the availability of basic knowledge, and, in part, the likelihood of discovery in support of proprietary therapeutic innovations.²⁰⁶ Another important, but less publicized, motivation for placing the EST data into the public domain was reputedly to preempt the patenting of these genetic sequences by private companies such as Incyte Pharmaceuticals and Human Genome Sciences.²⁰⁷

202. NIH GWAS Policy, *supra* note 200, at 49,296.

203. *See, e.g.*, ENCODE CONSORTIA DATA RELEASE, DATA USE, AND PUBLICATION POLICIES, *available at* <http://www.genome.gov/Pages/Research/ENCODE/ENCODEDAtaReleasePolicyFinal2008.pdf> (implementing a nine-month embargo period). *But see Human Microbiome Project (HMP): Overview*, THE NIH COMMON FUND, <http://nihroadmap.nih.gov/hmp/index.asp> (last visited Oct. 23, 2010) (implementing a data release policy for a recently initiated project to generate genomic sequence and related data for the thousands of microorganisms residing within the human body which contains no “embargo” period, but permitting data generators to withhold certain “analyses”—excluding sequence or related data—for a period of twelve months following generation).

204. *See* Toronto Authors, *supra* note 124.

205. *Id.* at 170.

206. *See* Press Release, Merck & Co., First Installment of Merck Gene Index Data Released to Public Databases: Cooperative Effort Promises to Speed Scientific Understanding of the Human Genome (Feb. 10, 1995), *available at* <http://www.bio.net/bionet/mm/bionews/1995-February/001794.html>.

207. *See* DON TAPSCOTT & ANTHONY D. WILLIAMS, WIKINOMICS: HOW MASS COLLABORATION CHANGES EVERYTHING 166 (2006); Marshall, *supra* note 168, at 1192; Rai, *supra* note 4, at 134 (describing the commercial strategy of Merck in releasing the EST data

A similar effort known as the SNP Consortium was formed in 1999 by a group of companies and the Wellcome Trust to identify and map genetic markers referred to as “single nucleotide polymorphisms” (SNPs) and to release the resulting data to the public domain, unencumbered by patents.²⁰⁸ The consortium accomplished this goal by filing U.S. patent applications covering the SNPs it discovered and mapped, and then ensuring that these applications were contributed to the public domain prior to issuance.²⁰⁹ This approach ensured that the consortium’s discoveries would act as prior art defeating subsequent third-party patent applications, with a priority date extending back to the initial filings. The SNP Consortium’s innovative “protective” patenting strategy has been cited as a model of private industry’s potential to contribute to the public genome commons.²¹⁰

Since the successful completion of the SNP Consortium project, numerous other privately funded research collaborations have adopted data release models that are similarly intended to place large quantities of genomic data into the public domain. In recent years, however, these efforts have included timing mechanisms in their data release policies. For example, the International SAE Consortium (SAEC) was formed in 2007 to fund research toward the identification of DNA markers for drug-induced serious adverse events. The SAEC adopted a “defensive” patent filing strategy similar to that of the SNP Consortium, which secures for data-generating scientists a period of exclusivity during which they have the sole ability to analyze data and prepare papers for publication.²¹¹ Like the policies adopted by government-

to the public); *see also* discussion of ESTs and the patenting debate surrounding them, *supra* note 188 and accompanying text. By 1998, the Merck Gene Index had released over 800,000 ESTs through GenBank. *See* TAPSCOTT & WILLIAMS, *supra*, at 166.

208. *See* TAPSCOTT & WILLIAMS, *supra* note 207, at 168 (noting consortium members’ concerns about biotech companies’ plans to patent SNPs and “sell them to the highest bidder”); Arthur L. Holden, *The SNP Consortium: Summary of a Private Consortium Effort to Develop an Applied Map of the Human Genome*, 32 *BIO TECHNIQUES* 22, 26 (2002) (“The overall IP objective is to maximize the number of SNPs that (i) enter the public domain at the earliest possible date, and, (ii) are free of third-party encumbrances such that the map can be used by all without financial or other IP obligations.”).

209. The SNP Consortium’s patenting strategy included filing patent applications covering all mapped SNPs and then converting those applications into statutory invention registrations (SIRs) or abandoning the applications after publication.

210. *See, e.g.*, Cook-Deegan & McCormack, *supra* note 176, at 217 (describing the consortium’s “unusual and sophisticated approach to keeping data in the public domain”); Marshall, *supra* note 168, at 1192 (noting the consortium’s “defensive move” deriving from the Merck Gene Index’s earlier strategy); Nunnally, *supra* note 128, at 252–53.

211. INT’L SAE CONSORTIUM LTD., DATA RELEASE AND INTELLECTUAL PROPERTY POLICY (last amended Nov. 5, 2009) (on file with author).

funded projects, the SAEC policy imposes a nine-month embargo on publication or presentation of publicly released data. But in addition the SAEC implements a delayed-release principle, allowing the data-generating group to retain data internally (i.e., to share it only among the SAEC participants) for a period of up to twelve months following data validation. Similarly, pharmaceutical manufacturer Pfizer and the non-profit Sage Bionetworks have recently entered into a collaboration in which Pfizer will reportedly make large amounts of its proprietary data available through Sage's "open access" mechanisms, but may retain data generated from the collaboration for a period of one year after the project is completed.²¹² The significance of these data "retention" approaches and their means of addressing policy objectives using a latency-based approach that differs significantly from the "embargo" approach will be discussed in Section V.A, *infra*.

B. SCIENTIFIC PUBLISHING

1. *Scientific Publishing and the Open-Access Debate*

The data release policies relevant to the genome commons focus on the release of experimental data prior to the publication of analytical results in a scientific journal. But an equally if not more contentious debate exists regarding the dissemination of scientific knowledge *after* it has been published. The debate centers on the accumulated scientific knowledge that is accessible only in proprietary scientific journals. According to one estimate, there were approximately 50,000 different scientific journals in print at the end of 2003,²¹³ many of which are published by commercial entities that charge significant subscription fees.²¹⁴ The proliferation of scientific journals coupled with high subscription costs has led to the cancelation of multiple journal subscriptions by academic libraries, often depriving researchers at

212. Stephen Strauss, *Pharma Embraces Open Source Models*, 28 NATURE BIOTECHNOLOGY 631, 632 (2010).

213. JOHN WILLINSKY, THE ACCESS PRINCIPLE: THE CASE FOR OPEN ACCESS TO RESEARCH AND SCHOLARSHIP 14 (2006) (citing Carol Tenopir).

214. According to one study, the average subscription cost of commercially published journals in the field of economics in 2001 was over \$1,600. Theodore C. Bergstrom, *Free Labor for Costly Journals?*, 15 J. ECON. PERSP. 183, 183 (2001). Specialist publications, particularly in the medical literature, can cost in the range of \$20,000 per year. Pamela Burdman, *A Quiet Revolt Puts Costly Journals on Web*, N.Y. TIMES, June 26, 2004, at B7 (citing the annual subscription rates of *The Journal of Comparative Neurology* (\$17,995) and *Brain Research* (\$21,269)).

those institutions of access to the current literature in their fields.²¹⁵ And because scientific publishers typically acquire the copyright in articles that they publish,²¹⁶ authors have traditionally been constrained in their ability to disseminate their published work outside of the publishers' established channels, whether in print or online.

Opposition to the influence that publishers have exerted over the dissemination of scientific data has led to a vocal and powerful "open access" movement among academic scholars and librarians.²¹⁷ They argue that the increasing control exerted by publishers over scientific results, and the decreasing number of scientists who can access these results in proprietary databases and publications, hinders the progress of science.²¹⁸ Accordingly, in 2003, a number of European academic and scientific institutions published the "Berlin Declaration," encouraging researchers to use open-access publishing methods in lieu of traditional subscription-based publication.²¹⁹ The largest private funders of biomedical research in the United States and Britain, the Howard Hughes Medical Institute and the Wellcome Trust, made a similar plea in the so-called "Bethesda Statement."²²⁰ Shortly thereafter, several major research institutions including Harvard University, the Massachusetts Institute of Technology, and University College London implemented policies requiring their faculty to deposit all research publications in open-access databases, even if previously published in a commercial journal.²²¹ This trend has since spread to institutions across the United States, Europe, Canada, Australia, and India.²²² The open-access

215. See WILLINSKY, *supra* note 213, at 24–25; NAS, RESEARCH DATA, *supra* note 82, at 78.

216. NAS, RESEARCH DATA, *supra* note 82, at 78.

217. "Open access" publishing typically connotes distribution through an online repository that is accessible to the public without charge. See generally Nancy Kranich, *Countering Enclosure: Reclaiming the Knowledge Commons*, in KNOWLEDGE AS A COMMONS, *supra* note 18 (explaining that the number of institutions globally that have adopted open-access publishing policies has doubled in the past year); WILLINSKY, *supra* note 213, at app. A; Peter Suber, *Creating an Intellectual Commons Through Open Access*, in KNOWLEDGE AS A COMMONS, *supra* note 18, at 172.

218. See sources cited *supra* note 217.

219. BERLIN DECLARATION ON OPEN ACCESS TO KNOWLEDGE IN THE SCIENCES AND HUMANITIES (2003), available at http://oa.mpg.de/files/2010/04/berlin_declaration.pdf.

220. Press Release, Peter Suber, Dir., Open Access Project, Bethesda Statement on Open Access Publishing (June 20, 2003), available at <http://www.earlham.edu/~peters/fos/bethesda.htm>.

221. See *Open-Access Publishing Gains Another Convert*, 459 NATURE 627 (2009); John Timmer, *MIT to Make All Faculty Publications Open Access*, ARS TECHNICA, Mar. 24, 2009.

222. NAS, RESEARCH DATA, *supra* note 82, at 79; Jocelyn Kaiser, *Free Journals Grow Amid Ongoing Debate*, 329 SCI. 896, 897 (2010) [hereinafter Kaiser, *Free Journals*].

directives adopted by these institutions directly contradict the publication policies of most scientific journals, and as such have led to difficult negotiations between publishers and universities adopting this approach. In many cases, publishers have reached a compromise with universities providing for a six- to twelve-month “embargo” period following publication before an article can be deposited in an open-access repository.²²³

2. *Open Access and Government-Funded Research*

An even more contentious debate exists surrounding the publication of government-funded research. According to one estimate, research funded by NIH alone results in approximately 60,000 scientific papers per year.²²⁴ Beginning in 2003, a growing number of scientists, archivists, and policy makers began to express the view that it was inappropriate for taxpayer-funded research to inure solely to the financial benefit of publishers, and that the public should have free access to taxpayer-funded research.²²⁵ Publishers, however, contend that federally funded research is separate from the subsequent publication of research results, and that the publication process, which encompasses both peer review and editorial oversight, is not subsidized by federal grant monies.²²⁶ They argue that “open access” publishing models, in which oversight and peer review may be less rigorous, permit less deserving research to be published.²²⁷

Nevertheless, in June 2004, the U.S. House Appropriations Committee instructed NIH to make all scientific publications generated by NIH-funded research available online without charge.²²⁸ The NIH implemented a voluntary “open access” policy in 2005, which was made mandatory in 2008.²²⁹ The current NIH policy provides that all publications arising out of

223. NAS, RESEARCH DATA, *supra* note 82, at 79.

224. WILLINSKY, *supra* note 213, at 2.

225. *Id.* It is useful here to differentiate between the information commons created by information accessible to the scientific community that subscribes to a given scientific journal and the broader public and scientific community. As noted by Reichman and Uhler, *supra* note 61, and others, the former community may be significantly smaller than the latter.

226. See Jeannie Baumann, *House Oversight Hearing Panelists Debate Open Access to Federally Funded Research*, BNA BIOTECH WATCH, Aug. 4, 2010 (citing Allan Adler of the Association of American Publishers).

227. Kaiser, *Free Journals*, *supra* note 222, at 896–97.

228. See Suber, *supra* note 217, at 203 n.41.

229. Nat'l Insts. of Health, Revised Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research, NOT-OD-08-033 (Apr. 7, 2008), available at <http://grants.nih.gov/grants/guide/notice-files/not-od-08-033.html> (implementing Consolidated Appropriations Act, 2008, Pub. L. No. 110-161, § 218, 121 Stat. 1844, 2187 (2007)).

NIH-funded research data must be submitted to the National Library of Medicine's publicly accessible PubMed Central digital archive within one year following publication.²³⁰ The one-year period established by the NIH policy provides the publishing industry with a period of exclusive article access, to avoid the loss of paid subscription revenue.²³¹ This period reflects a compromise among publishers who opposed the measure in general, NIH, which initially proposed a six-month period, and open-access advocates, who argued for immediate public release of articles.²³² Funding agencies in Europe have enacted mandates similar to the NIH policy.²³³ Legislation recently proposed in the U.S. Congress would expand the NIH's open-access mandate to all federal agencies with a shortened embargo period of six months.²³⁴ Not surprisingly, legislation supported by the publishing industry and opposing both initiatives has also been proposed in Congress.²³⁵

3. *Alternative Open-Access Models*

The scientific publishing industry has not, however, uniformly opposed open-access initiatives and a few publishers have even embraced them. Journals such as the *New England Journal of Medicine (NEJM)* and *Molecular Biology of the Cell (MBC)* now voluntarily make their contents publicly available after a waiting period (six months in the case of *NEJM*; two months in the case of *MBC*).²³⁶ *NEJM* and *MBC*, however, are both published by scientific trade associations,²³⁷ and thus serve their members through multiple channels, of which journal publication is only one. The largest publishers of scientific journals, Reed Elsevier (approximately 1,800 titles), Taylor and Francis (more than 1,000 titles), and Springer Verlag (more than 500 titles), which collectively control sixty percent of scientific research content, are

230. It has been estimated that, as of April 2010, PubMed Central contains two million scholarly articles. Declan Butler, *US Seeks to Make Science Free for All*, 464 NATURE 822, 822 (2010).

231. *Id.*

232. WILLINSKY, *supra* note 213, at 3; Editorial, *Open Sesame*, 464 NATURE 813 (2010).

233. Butler, *supra* note 230, at 822.

234. Federal Research Public Access Act (FRPAA), H.R. 5037, 111th Cong. (2010).

235. Fair Copyright in Research Works Act, H.R. 801, 111th Cong. (2009) (would prohibit federal agencies from adopting open-access publication policies). For a discussion of this legislative proposal, see, for example, James Boyle, *Misunderestimating Open Science*, FIN. TIMES, Feb. 24, 2009. A representative of publisher Reed Elsevier recently called the proposed open-access requirements "a means for facilitating international piracy." Butler, *supra* note 230, at 822.

236. WILLINSKY, *supra* note 213, at 68.

237. *NEJM* is published by the Massachusetts Medical Society and *MBC* by the American Society for Cell Biology.

commercial interests with significant subscription and reprint revenues at stake.²³⁸

One open-access approach that could potentially satisfy the revenue requirements of commercial publishers is the so-called “author pays” model, in which articles are made freely accessible to readers, but the journal charges authors a fee to cover editorial, publishing, and peer review costs.²³⁹ The author-pays model was pioneered by the non-profit Public Library of Science (PLOS), which operates seven journals covering biomedical topics such as genetics, biology, and medicine, and which was co-founded by Nobel laureate and former NIH Director Harold Varmus.²⁴⁰ Although the author-pays model has still not made significant inroads in most scientific disciplines, PLOS’s flagship journal *PLoS ONE* has seen significant increases in submissions since its foundation in 2006.²⁴¹ Publication fees for PLOS journals range from \$1,350 (for *PLoS ONE*) to \$2,900 (*PLoS Biology* and *PLoS Medicine*),²⁴² and an increasing number of funding agencies and foundations appear to be willing to pay open-access publication fees for research that they fund.²⁴³

Commercial publishers, perhaps sensing the prevailing wind in support of open access, have also taken tentative steps toward open-access

238. WILLINSKY, *supra* note 213, at 18.

239. This approach should be distinguished from the customary practice of many professional societies (such as the American Physiological Society) to charge authors publication fees in order to defray member subscription costs. See Dale J. Benos, L. Gabriel Navar & Margaret Reich, *Publishing in the Journals of the APS: Why are Authors Charged Fees?*, 278 AM. J. PHYSIOLOGY GASTROINTESTINAL LIVER PHYSIOLOGY 663, 663 (2000) (“Like many other association publishers, APS is able to keep subscription prices low by sharing some of the cost of publishing the journals with the authors who submit manuscripts. . . . Many commercial publishers do not charge authors for publication (i.e., page charges) but have much higher subscription prices.”); see also *Information for Authors*, PROCEEDINGS NAT’L ACAD. SCI. (Oct. 2010), <http://www.pnas.org/site/misc/iforc.shtml> (detailing fees per page and per color figure); *Instructions to Authors*, J. VIROLOGY, http://jvi.asm.org/misc/journal-ita_pub.dtl (last updated Sept. 2010) (detailing fees per page and per color figure).

240. See PUB. LIBR. SCI., <http://www.plos.org> (last visited Nov. 8, 2010); Kaiser, *Free Journals*, *supra* note 222, at 897; Kranich, *supra* note 217, at 96.

241. See Butler, *supra* note 230, at 823.

242. *Publication Fees for PLoS Journals*, PUB. LIBR. OF SCI., <http://www.plos.org/journals/pubfees.php> (last visited July 4, 2010).

243. Butler, *supra* note 230, at 823. A group of major research universities including Harvard, M.I.T., Dartmouth, Cornell, Berkeley, Columbia, Memorial Sloan-Kettering, and the University of Ottawa have formed a group called the Compact for Open-Access Publishing Equity to advocate for greater payment of open-access publication fees by research funders. COMPACT FOR OPEN-ACCESS PUBLISHING EQUITY, <http://www.oecompact.org/compact/> (last visited Oct. 23, 2010).

publishing. In 2008, for example, the German publisher Springer Verlag acquired BioMed Central, which publishes more than 200 peer-reviewed open-access journals.²⁴⁴ Other commercial publishers, such as Nature Publishing Group, have experimented with a hybrid approach under which authors have the option either to publish in the traditional manner or to pay to make individual articles available on an open access basis.²⁴⁵ Nevertheless, the largest scientific publishers continue to oppose legislation and policies designed to mandate open access for government-funded research.²⁴⁶

V. APPLYING THE LATENCY FRAMEWORK

The two case studies presented in Part IV illustrate well-known examples of the construction of science commons. I have chosen to highlight them because they exemplify the use of temporal features by policy designers to achieve specific policy outcomes in a private ordering context and to reach compromise among stakeholders with competing interests. In this Part, I will apply the latency analysis introduced in Part III to these case studies.

A. LATENCY IN THE GENOME COMMONS

1. *Stakeholders and Policy Design Considerations*

The primary stakeholders engaged in the policy debate over data release in the genome commons are funding agencies (primarily the U.S. NIH and the Wellcome Trust in the UK), data generators, and data users.²⁴⁷ More recently, the critical role of scientific journals has been recognized, particularly with respect to the publication of articles that may be in violation of embargo requirements and the need to offer meaningful and career-enhancing publication opportunities to data generators.²⁴⁸ While data release policies are typically drafted and adopted by funding agencies, NIH in particular has given substantial deference to the views and opinions of the scientific community when developing policy, while also seeking to represent

244. See BIOMED CENT.: THE OPEN ACCESS PUBLISHER, <http://www.biomedcentral.com> (last visited Nov. 8, 2010).

245. See Press Release, Nature Publ'g Grp., Open Access Options on 7 More Nature Publishing Group Journals (May 19, 2010), available at http://www.nature.com/press_releases/openaccess.html.

246. See Baumann, *supra* note 226.

247. See Ft. Lauderdale Principles, *supra* note 189, at 2–3 (referring to the “tripartite” sharing of responsibility among these groups). Because most large-scale genomic databases are operated by governmental agencies (e.g., NIH’s GenBank and dbGaP), the interests of such data intermediaries are typically aligned with those of governmental funders.

248. See *id.* at 4; Toronto Authors, *supra* note 124, at 170.

the interests of the general public.²⁴⁹ Thus, the role and influence of other stakeholder groups is not to be underestimated: the development of data release policies in the genome sciences has been a process of negotiation and compromise.

The evolution of the genome commons illuminates three principal policy considerations that motivated these stakeholder groups: (1) promoting the advancement of science by making genomic data as widely available as possible (“scientific advancement,” typically espoused by funders and public-interest advocates); (2) giving data generators priority in the publication of papers based on their data (“publication priority,” typically espoused by data generators); and (3) minimizing patent-related encumbrances on genomic data sets (“minimizing encumbrances,” espoused by both funders and data users).²⁵⁰ The interplay of these three design considerations, and the latency-based compromises that were effected to satisfy competing requirements of relevant stakeholders, resulted in the policies that are in effect today.

249. See, e.g., Collins et al., *supra* note 178, at 846 (2003) (describing community involvement in setting goals for national genomics research program). NIH has most recently requested public comment and feedback on data release policies in October 2009. Press Release, Nat'l Insts. of Health, Notice on Development of Data Sharing Policy for Sequence and Related Genomic Data, NOT-HG-10-006 (Oct. 19, 2009), available at <http://grants.nih.gov/grants/guide/notice-files/NOT-HG-10-006.html>.

250. Other factors such as protection of human subject data also played a substantial role in the development of genomic data release policies. See *supra* note 113 and accompanying text. These factors, however, have remained relatively invariant and not subject to compromise, and are thus not addressed in detail here.

Table 1: Latency Characteristics of Genomics Data Release Policies

	Knowledge Latency (KL)	Rights Latency (RL)
U.S. Federally Funded Genomics Policies²⁵¹		
NIH-DOE 1992 Guidelines ²⁵²	6 months	0
Bermuda Principles (1996) ²⁵³	24 hours	0
NIH 1996 Policy ²⁵⁴	Short ²⁵⁵	0
NIH 1997 Policy ²⁵⁶	24 hours	0
NIH 2000 Policy ²⁵⁷	24 hours–7 days	0
NIH 2003 Policy ²⁵⁸	24 hours	0
Intl. HapMap Project (2003) ²⁵⁹	Short	0
ENCODE Pilot (2003) ²⁶⁰	Short	0
GAIN (2006) ²⁶¹	Short	9 months (generally)
NIH GWAS Policy (2007) ²⁶²	Short	12 months
ENCODE + modENCODE (2008) ²⁶³	Short	9 months
Non-Federal Genomics Policies		
SNP Consortium (1998) ²⁶⁴	30/90 days	0
Intl. SAE Consortium (2007) ²⁶⁵	12 months	9 months

251. As discussed *supra* Section IV.A.1, many of the NIH and other policies phrase timing requirements as encouragements or suggestions. While the enforceability of such “soft” requirements is open to question, for purposes of this analysis, the timing features of these policies are listed notwithstanding potential issues with enforcement.

252. *See supra* note 170.

253. *See supra* note 35.

254. *See supra* note 188.

255. A number of policies include requirements that data be released as rapidly as possible following validation or the satisfaction of some other initial condition. This requirement can be translated into a numerical value depending on the particular type of data involved and customary practice in the field. For purposes of this Article, however, these policies should simply be viewed as having quite short knowledge latencies, probably on the order of one month or less.

256. *See supra* note 172 and accompanying text.

257. *NHGRI Policy for Release and Database Deposition of Sequence Data*, NAT’L HUM. GENOME RES. INST. (Dec. 21, 2000), <http://www.genome.gov/page.cfm?pageID=10000910>.

258. *Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-Scale Sequencing and Other Community Resource Projects*, NAT’L HUM. GENOME RES. INST. (Feb. 2003), <http://www.genome.gov/10506537>.

259. The Int’l HapMap Consortium, *The International HapMap Project*, 426 NATURE 789, 793 (2003) (“All data . . . will be released rapidly.”).

260. *ENCODE Project Data Release Policy (2003–2007)*, NAT’L HUM. GENOME RES. INST., <http://www.genome.gov/12513440> (last updated Apr. 20, 2010).

261. *See supra* notes 198–201.

262. *See supra* note 200.

263. *ENCODE CONSORTIA DATA RELEASE, DATA USE, AND PUBLICATION POLICIES*, <http://www.genome.gov/Pages/Research/ENCODE/ENCODEDataReleasePolicyFinal2008.pdf> (last visited Oct. 23, 2010).

264. *See supra* notes 208–09.

265. *See supra* note 211.

Pfizer – SAGE (2010) ²⁶⁶	12 months	0
-------------------------------------	-----------	---

2. Latency as a Policy Design Tool—Retention versus Embargo

As demonstrated by Table 1, in every NIH data release policy following the adoption of the Bermuda Principles, genomic data must be released rapidly to public databases. The twofold policy motivations for this requirement have been discussed above: there is the explicit motivation of accelerating the progress of science, and a less explicit, but strongly implied, motivation of limiting patent encumbrances on genomic data. However, once concerns regarding publication priority between data generators and data users arose, there was a need for policy change. Most other federal agencies and private initiatives in the genome sciences have addressed this conflict by permitting data generators to withhold their data from the public for a specified time period, generally in the range of twelve months,²⁶⁷ after which it would be released unencumbered (this can be called a “retention strategy”). In terms of latency analysis:

Retention Strategy

knowledge latency (KL) = 12 months

rights latency (RL) = 0

NIH policy makers, however, elected to take an alternative approach. Instead of allowing data generators to retain their data for a protected period and then releasing it unencumbered, each of the new NIH policies continues to require rapid data release, but imposes a publication embargo on users (an “embargo strategy”). Thus, in the case of the GAIN NIH GWAS Policy and other recent NIH policies:

Embargo Strategy

knowledge latency (KL) = 0/short

rights latency (RL) = 12 months

These two strategies are summarized in Table 2.

266. See *supra* note 212.

267. See *supra* note 173 and accompanying text.

Table 2: Latency Analysis of Genomic Data Release Policies

	Knowledge Latency (KL)	Rights Latency (RL)
Bermuda	~0	0
Embargo (GAIN/GWAS)	~0	12 mo.
Retention (SAGE, SAEC)	12 mo.	0+

On paper, the result for data generators appears to be the same under both an embargo and a retention strategy: in either case they have a period of twelve months during which they retain exclusive rights to publish papers analyzing the data. But in practice there are material differences between the two strategies.²⁶⁸ These differences are driven by material externalities that distinguish government-funded projects from privately funded projects.

A retention strategy lengthens knowledge latency and, by definition, extends the time before data is released to the public. NIH has repeatedly stated that genomic data should be released as rapidly as possible for the advancement of science and the public good.²⁶⁹ The embargo approach accomplishes this goal by minimizing knowledge latency while still protecting the data generators' publication interests. However, the embargo strategy involves a significant tradeoff in terms of effectiveness. Usage embargos in NIH's recent data release policies are typically embodied in click-wrap agreements or online certifications that must be made with a data request.²⁷⁰ A discussion of the enforceability of these mechanisms (whether in strictly legal terms or in terms of the ability and willingness of journals and funders to police compliance) is beyond the scope of this Article.²⁷¹ However, it is self-evident that even the *most* robust contractual embargo provides the data generator with *less* protection than withholding the data from the public (i.e., if the user has no data, he or she cannot breach his or her obligation to

268. The distinction between these two approaches is illustrated *supra* Figures 2 and 3.

269. See *supra* note 178 and accompanying text.

270. A "click-wrap" agreement (alternatively referred to as a "click-through" or "click-to-accept" agreement or license) is "an electronic form agreement to which [a] party may assent by clicking an icon or a button or by typing in a set of specified words." Christina L. Kunz et al., *Click-Through Agreements: Strategies for Avoiding Disputes on Validity of Assent*, 57 BUS. LAW. 401, 401 (2001).

271. For a general discussion of the enforceability of click-wrap agreements, see GEORGE B. DELTA & JEFFREY H. MATSUURA, LAW OF THE INTERNET, § 10.05 (2d ed. Supp. 2010–2012). Rebecca Eisenberg, who analogizes click-wrap agreements for genomics projects to the open source software General Public License (GPL), raises questions about the enforceability of such agreements. Eisenberg, *supra* note 101, at 1028. In addition to legal enforceability, serious issues of detection and policing arise when considering the efficacy of obligations imposed in this manner.

refrain from publishing). Moreover, a retention strategy gives the data generator a true head start with respect to the data, during which no third party may analyze or build upon it (witness the seventeen-year holdback of *Ardi* data). In contrast, an embargo strategy enables third parties both to analyze and to build upon the data *during* the embargo period, putting them in a position to submit papers for publication the moment the embargo expires. Finally, under a rapid-release embargo system, data generators run the risk that journals less familiar with rapid pre-publication data release will consider the initial release of data in a public database to constitute a prior publication that would disqualify the data generator from subsequently publishing his or her analysis.²⁷²

With all of these comparative disadvantages to the data generator, why do the NIH policies adopted since 2006 emphasize rights latency over knowledge latency? The answer may lie in NIH's approach to intellectual property. As discussed above, NIH must operate within the constraints of the Bayh-Dole Act. Thus, while NIH's various post-Bermuda data release policies all acknowledge the requirements of the Bayh-Dole Act, they evidence a general bias against patent encumbrances on genomic data.²⁷³ The enforceability, however, of policy provisions that merely "urge" or "encourage" data generators and users not to seek patents on inappropriate subject matter is open to some doubt.²⁷⁴ Lacking a strong policy tool with which to limit expressly the patenting of genomic information, NIH policy makers have employed rapid pre-publication data release requirements as a surrogate for achieving the same result.²⁷⁵ The Bermuda Principles, in

272. Most scientific journals will only accept articles for publication if they have not previously been published or released. While journals specializing in genomics and allied fields, or which are otherwise actively engaged in the conversation regarding pre-publication data release, would ordinarily not consider pre-publication data release to constitute a prior publication for this purpose, the author has encountered anecdotal evidence suggesting that journals that are less familiar with the practice have rejected articles by data generators on this basis.

273. See, e.g., *NHGRI 1996 Policy*, discussed *supra* notes 187–88 and accompanying text; *ENCODE Pilot Policy*, discussed *supra* note 203; *NIH GWAS Policy* discussed *supra* Section IV.A.1.e.

274. See Rai & Eisenberg, *supra* note 187, at 309.

275. An interesting early exception is the International HapMap Project, which was organized in 2002 by scientists and funding agencies from Japan, the United Kingdom, Canada, China, Nigeria, and the United States to develop a "haplotype" map of the human genome. See The Int'l HapMap Consortium, *supra* note 259. Data generated by the project were released "rapidly" into publicly accessible databases, but a click-wrap agreement that purportedly prohibited data users from filing patent applications on project results was also used. *Id.* at 793. Though this precise history is somewhat unclear, Rebecca Eisenberg reports

particular, and their adoption by NIH in 1997 and reaffirmation in 2003, ensured that genomic data from the HGP and other large-scale sequencing projects would be made publicly available before data generators had an opportunity to file patent applications covering any “inventions” arising from that data, and in a manner that ensured its availability as prior art against third party patent filings at the earliest possible date.²⁷⁶

In contrast, non-governmental groups such as SAEC adopted lengthier knowledge latency periods to protect the publication priority of their researchers more effectively, but they did so in conjunction with explicit patent-defeating strategies.²⁷⁷ These groups, unlike federal agencies, have the freedom to impose express contractual limitations on patenting and to adapt defensive patenting strategies without running afoul of the requirements of the Bayh-Dole Act. Thus these policies may be most optimal with respect to the policy goals of minimizing encumbrances and protecting data generators’ publication priority, but are less optimal than the government-led policies in terms of broad disclosure of knowledge and scientific advancement. The trade-offs may reflect the different policy objectives of government NIH funders versus private commercial-consortium organizers. Table 3 summarizes the differing policy outcomes achieved by adjustment of latency variables in the genome commons.

that a formal Data Release Policy was adopted by the project as late as 2004 and was intended to supersede the click-wrap structure. Eisenberg, *supra* note 101, at 1026. It is possible that this policy was implemented in response to objections by federal funding agencies to the earlier contractual patenting prohibitions.

276. See discussion *supra* notes 177–82 and accompanying text.

277. See discussion *supra* notes 210–11 and accompanying text.

Table 3: Policy Outcomes in of Genomic Data Release Policies

	Scientific Advancement (SA) (1)	Minimizes Encumbrances (ME) (2)	Publication Priority (PP) (3)
Bermuda (KL=0/RL=0)	High	Medium	Low
Embargo (KL=0/RL=1)	High	Medium	Medium
Retention (KL=1/RL=0)	Medium	High	High

- (1) Scientific Advancement (SA) is highest under policies that minimize knowledge latency (KL). While data is retained by data generators in private, overall scientific advancement cannot occur. Thus, under a retention strategy, SA is classified as “Medium” (as opposed to “Low,” given that data *is* released to the public, albeit after a waiting period).
- (2) Minimization of encumbrances (e.g., patent protection on data) is achieved by two different means. Private groups implementing retention strategies have employed contractual anti-patenting policies coupled with “protective” patent filings, resulting in a high degree of freedom from patent encumbrances. Government-funded projects, which cannot avail themselves of these techniques, are left with “hortatory” encouragements against patenting, coupled with early disclosure of information as prior art (low KL). While the effectiveness of these measures is debatable, they are likely not as strong as the approach taken under the retention strategy, hence a “Medium” rating.
- (3) Publication priority (PP) for data generators was explicitly sacrificed under public projects such as the HGP in the service of scientific advancement. While embargo policies have attempted to improve priority for data generators, the enforceability of contractual embargo provisions is less certain than simple withholding of information under a retention policy.

B. LATENCY AND SCIENTIFIC PUBLISHING

1. Stakeholders and Policy Design Considerations

In contrast to the policy negotiations over the genome commons, which, despite their far-reaching implications for biomedical research, directly affect a relatively discrete community of scientists and funders, the debate over scientific publishing implicates all scientific disciplines and many other fields of scholarly research including the humanities and arts. As such, the community of stakeholders relevant to this debate is large and varied. Nevertheless, the stakeholder categories identified in Section III.A.3 apply to this analysis as well.

As in every information commons, data generators play a central role. In this context, data generators are the scientists who have an interest in publishing their results for purposes of career and reputational enhancement

(i.e., the vast majority of scientists).²⁷⁸ Unlike the genome commons, in which the communities of data generators and data users are more or less distinct (though converging as the cost of genomic sequencing continues to decrease), data-generating scientists in the context of scientific publishing are nearly always data users as well. Thus, in addition to the need to publish their own work, scientists must also access and utilize the published work of others. Research institutions are typically responsible for ensuring that their scientists have access to sufficient information resources, typically through their library function. The library community has been particularly vocal in the scientific publishing debate, and has decried rising subscription costs, publishing practices, and legal regimes that have led to so-called “enclosures” of the science commons and the inaccessibility of current scientific information to researchers.²⁷⁹ They argue that the advancement of science is imperiled by this enclosure and the decreasing number of individuals having access to potentially important scientific discoveries and data.²⁸⁰ This view has also been expressed by government funding agencies that argue for greater public access to taxpayer-supported research. As in the case of the genome commons, NIH has taken a leading role in representing the public interest in this debate.²⁸¹ Publishers—commercial publishers in particular—are on the opposite side of this debate and have argued that the services they provide in terms of selection, peer review, editing, and distribution are costly; that financial incentives must exist if scientific publication is to continue; and that the movement toward open-access publication is likely to result in scientific articles that are less reliable and of lesser quality.²⁸² Thus, where their principal policy objective is profit-making (or, in more sympathetic terms, financial survival), publishers also invoke a policy argument sounding in scientific advancement.

2. *Compromises Effected Through Latency Choices*

While the differences among stakeholders described in the preceding section may seem insurmountable at first glance, these differences have been addressed, at least for the moment, through a series of temporal policy

278. See *supra* Section III.A.1.

279. See, e.g., Kranich, *supra* note 217, at 87–107.

280. Empirical evidence supporting this claim, however, is still somewhat sparse and inconclusive. See Kaiser, *Free Journals*, *supra* note 222, at 898.

281. See *supra* notes 228–35 and accompanying text.

282. See Editorial, *supra* note 232, at 813. By one estimate, publication costs for a single article in the most prestigious scientific journals can run up to \$10,000. Kaiser, *Free Journals*, *supra* note 222, at 897.

compromises that adjust knowledge latency²⁸³ in the relevant science commons. To recapitulate: (1) universities and publishers have negotiated limited exclusivity periods of *six-to-twelve months* before university researchers are required to release their work to the public, (2) membership organizations that publish scientific journals such as the New England Journal of Medicine and Molecular Biology of the Cell, in response to member pressures, voluntarily permit open-access release of articles following an exclusivity period of up to *six months*, (3) NIH has mandated that all publications arising from NIH-funded research be released to the PubMed Central database *one year* after publication, and (4) additional legislation that has been proposed in Congress would extend the NIH mandate to all federal agencies and reduce the holding period to *six months*. Table 4 summarizes these compromises in terms of the exclusivity periods granted to publishers.

Table 4: Latency-Based Compromises in Scientific Publishing

Policy	Exclusive Period (knowledge latency)
Publisher-University Negotiated Embargo Periods	6–12 months
New England Journal of Medicine (NEJM)	6 months
Molecular Biology of the Cell (MBC)	2 months
NIH Open Access Policy (2008)	12 months
Proposed Federal Research Public Access Act (2010)	6 months

In each of these policies, scientific publishing is conducted in its current form and the publisher enjoys a limited-duration period of exclusivity. Following this period of exclusivity, release of the published work via open-access channels is authorized. It is significant that the policies illustrated above were developed through a wide range of processes, from unilateral adoption (in the case of NEJM and MBC), to bilateral negotiation (in the case of publisher-university embargos), to executive agency action (in the case of the NIH policy), to legislative action (in the case of the proposed FRPAA).

It is also significant that publishers, scientists, government, and research institutions have seemingly accepted the adjustment of knowledge latency as a suitable vehicle for addressing the differing concerns and objectives of

283. Because no substantial limitations on the use of data contained in scientific articles are imposed by journals (other than typical prohibitions on reproduction existing under copyright law), the adjustment of rights latency has not played a significant role in the debate over scientific publishing. Of course, information published in journals may be protected by patents, but patenting is currently not addressed by journal publication policies.

these diverse stakeholder groups. Though positions regarding the desired length of holding or embargo periods still differ, it appears that the scientific publishing community is converging on a period in the range of six to twelve months as an appropriate compromise.²⁸⁴ This convergence resembles that observed with respect to the genome commons. In both instances, competing stakeholder groups have addressed seemingly intractable policy differences through the adjustment of latency variables.

C. SYNTHESIS AND BROADER APPLICATIONS

Three conclusions about the design of scientific information commons can be drawn from the above case studies. First, these examples demonstrate that multiple stakeholder groups with competing and divergent interests can, and do, achieve compromise when constructing information commons. Second, once these commons are constructed, they can and do achieve desired social benefits. Third, the resolution of many of the conflicts among stakeholder groups in constructing information commons can be addressed through the adjustment of latency variables, whether knowledge latency and rights latency in tandem (as in the genome commons), or knowledge latency alone (as in scientific publishing). Why is this the case? Why do latency variables, among the many other characteristics of information commons,²⁸⁵ exert such a strong effect on the outcome of rules negotiations, whether privately ordered or governmental, and why are they the variables, among so many others, that have attracted the attention of policy makers and other stakeholder groups?

Several potential explanations present themselves. First, latency variables are fundamental to information commons, as they define the rate at which knowledge enters, and becomes usable within, the commons. Just as the rate of depletion of a physical resource commons such as a forest, pasture, or fishery is critical to understanding and preserving such a resource, latency is critical to understanding and building a commons of information. As such, latency should play a central role in rulemaking and design with respect to

284. To be sure, some publishers still argue that a twelve-month “embargo” period is too short to enable them to recoup their publication costs and that no single time period will be uniformly acceptable. Baumann, *supra* note 226.

285. Arun Agrawal has identified twenty-four separate variables that have been considered in analyses of common pool resources. Arun Agrawal, *Sustainable Governance of Common-Pool Resources: Context, Methods, and Politics*, 32 ANN. REV. ANTHROPOLOGY 243, 254 (2003). In his response to Madison, Frischmann, and Strandburg, *supra* note 23, Gregg Macey also points to the “many-variable” problem as a weakness of functionalist accounts of commons structures. Gregg P. Macey, *Cooperative Institutions in Cultural Commons*, 95 CORNELL L. REV. 757, 768–69 (2010).

information commons and, at a minimum, should be considered in the analysis of common information resources, whether under the IAD framework, the revised framework of Madison, Frischmann, and Strandburg, or future analytical frameworks.

Second, latency variables are structurally suited to effect compromise among competing interests. They are readily quantifiable, infinitely divisible, and intuitively understandable to experts and non-experts alike.²⁸⁶ Unlike other commons characteristics that could potentially be used to achieve similar policy goals (e.g., scientific advancement, publication priority, and minimization of encumbrances) such as, for example, the limited scope of intellectual property protection and the exemptions for “fair use” and experimental use,²⁸⁷ the adjustment of well-defined temporal periods is both efficient and definite.

Third, latency variables implicitly establish that the *time* during which information resides within (or is usable within) a commons and, conversely, the time during which information is held outside of the commons (or is subject to usage limitations), has a quantifiable *value*.²⁸⁸ When a value for a commodity exists, parties can negotiate based on their own economic assessments and, as has been shown in the above case studies, strike a bargain. In this sense, latency variables are satisfyingly reductionist. They enable stakeholders to distil complex policy positions into simple numerical values that are susceptible to numerical compromise, and to negotiate on the

286. Negotiation over timing begins at an early age, from requests to postpone bedtime to the length of periods allotted for play, naps, and homework.

287. These concepts have proven to be notoriously vague and difficult foundations on which to base behavior.

288. Interestingly, the *value* of commons-based information has not been widely studied. In contrast is the value of encumbered information, notably that covered by patents and copyrights, which has been studied extensively. See, e.g., Gerald J. Mossinghoff, *Overview of the Hatch-Waxman Act and Its Impact on the Drug Development Process*, 54 FOOD & DRUG L.J. 187 (1999) (describing the compromises and motivations involved in crafting the Hatch-Waxman Act, which included patent term extensions for pioneer drugs undergoing long FDA approval processes); Avishalom Tor & Dotan Oliar, *Incentives to Create Under a “Lifetime-Plus-Years” Copyright Duration: Lessons from a Behavioral Economic Analysis for Eldred v. Ashcroft*, 36 LOY. L.A. L. REV. 437, 439–40 (2002) (arguing that the value created by copyright term extensions are not always offset by incentives to produce copyrighted works); Sabra Chartrand, *Patents; Congress Has Extended Its Protection for Goofy, Gershwin and Some Moguls of the Internet*, N.Y. TIMES, Oct. 19, 1998, at C2 (noting that the Copyright Term Extension Act of 1998 was passed with “intense lobbying for it from powerful copyright holders”); Andrew Pollack, *A Company Races to Keep a Drug Patent*, N.Y. TIMES, Mar. 19, 2010, at B1 (noting the large value of a patent extension for a drug patent, loss of which would “cost[] the company hundreds of millions of dollars in lost sales”).

basis of these ostensibly objective criteria rather than more value-laden positions such as preferences for intellectual property protection versus the public domain. The introduction of latency variables into the commons formation discourse thus enables parties to sidestep (usually fruitless) debate over entrenched values and reach agreement based on simple numerical calculation.²⁸⁹ Consequently, the use of latency variables in policy design helps to reduce negotiation transaction costs, those perennial hobgoblins of economic analysis, and hopefully to achieve efficient and equitable results for all stakeholders.²⁹⁰ And, perhaps more importantly, the achievement of agreement on the basis of latency-based compromise may enable socially beneficial information commons to be formed where otherwise they would not, or where their formation otherwise would be substantially delayed due to negotiation roadblocks.

Finally, one cannot help but speculate regarding the significance of the convergence of latency variables in the two case studies presented in this Article. In the example of the genome commons, both the “retention” approach (via knowledge latency) and the “embargo” approach (via rights latency) have resulted in twelve-month waiting periods before data is freely

289. See *supra* note 29. Roger Pielke divides political conflicts regarding science-based policy into two broad categories: “tornado politics” (conflicts that can be resolved through the application of more and better scientific data, such as the decision regarding the most prudent course of action in the face of an approaching tornado) and “abortion politics” (conflicts that are characterized by differing values systems that are unlikely to be resolved through the application of additional data, such as the debate regarding the legalization of abortion). ROGER A. PIELKE, JR., *THE HONEST BROKER: MAKING SENSE OF SCIENCE IN POLICY AND POLITICS* 40–44 (2007). The use of latency variables as surrogates for value-laden positions regarding, for example, openness and intellectual property, transforms the debate regarding information commons to one that more closely resembles Pielke’s “tornado politics” than “abortion politics.”

290. This is not to say, of course, that the reduction of deeply held values to numerical surrogates is necessarily a desirable approach in all circumstances. In the case of commons formation, however, this technique has proven to be a useful tool in overcoming negotiation roadblocks and ensuring that commons are created. It is also worth noting that the “negotiation” strategies enabled by the use of latency analysis may only be effective for stakeholder groups that are adequately represented at the negotiating table. That is, when stakeholder groups are wholly or partially disenfranchised or underrepresented (in the case, for example, of citizens of developing countries), it may be difficult for such stakeholders to represent their collective position effectively, with or without the availability of latency techniques. Fortunately, in the cases examined in this article and most other cases involving commons of scientific information, governmental policy makers have typically sought to represent the interests of the public at large. See discussion *supra* note 113 and accompanying text.

usable in the commons.²⁹¹ In the case of scientific publishing, though negotiations and debate are ongoing, it appears that a six- to twelve-month “exclusivity” period may emerge as a satisfactory resolution for most parties. It is tempting to speculate whether these real-world latency periods represent, or at least begin to approach, values that may be deemed to be “efficient” or “optimal,” and whether the final values of these latency periods can tell us something about the intrinsic value of scientific knowledge in the market economy. But even if one postulated that an optimal latency value may exist for a particular scientific commons, it is likely that this value would be difficult, if not impossible, to translate to other scientific disciplines, as norms, practices, and expectations are likely to vary among scientific disciplines that are as diverse as, for example, genomics, astrophysics, botany, and paleoanthropology (as demonstrated by the case of *Ardi*). It is also likely that the balance between knowledge latency and rights latency, and the efficiencies achieved by adjusting these twin variables, will vary according to the particulars of the scientific discipline and the commons being formed, just as the approaches and policy considerations that arose in the case studies of genomics and open-access publishing differ.²⁹² I am hopeful that further

291. Upon observing this result, one is drawn to Coase’s famous theorem which postulates that, absent transaction costs, the final allocation of resources among transacting parties will be the same, no matter how legal rules may define initial entitlements. Ronald Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960). Shubha Ghosh invokes Coase’s theorem in discussing the formation of public commons of copyrighted material, explaining that it may be “largely irrelevant who has the legal rights in the creation (the creator or the public at large) as long as bargaining is allowed.” Ghosh, *supra* note 155, at 222–23. Thus, in Ghosh’s estimation, information commons will be created, whether legal rules impose protective intellectual property or public domain regimes. In the case of the genome commons, one is tempted to wonder whether it is meaningful that twelve-month data waiting periods have arisen in two different contexts: that of private consortia, which explicitly prevent patenting of results, and that of NIH-funded programs, in which patenting cannot legally be prohibited. Though further study and economic analysis are required, such an outcome suggests itself as a natural proof of Coase’s theorem in the spirit of Robert Ellickson’s landmark study of rancher-landowner interactions in Shasta County, California. ROBERT C. ELLICKSON, *ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES* 2–3 (1991).

292. One extreme example of the importance of temporal factors to an emerging information commons can be found in nineteenth century attempts to extend copyright protection to published news and, in particular, the proposed 1884 Bill Granting Copyright to Newspapers. S. 1728, 48th Cong. (1884). The so-called News Copyright Bill reflected a realization by the Associated Press and other metropolitan news sources that significant economic value arose from being the first to print the news. See Robert F. Brauneis, *The Debate Over Copyright in News and its Effect on Originality Doctrine*, in *INTELLECTUAL PROPERTY PROTECTION OF FACT-BASED WORKS* 39, 54–56 (Robert F. Brauneis, ed. 2009). Accordingly, the Bill sought to extend to newspaper publishers an eight-hour exclusive right

application of the latency analysis presented herein to additional commons of scientific information may shed further light on the complex economic dynamics of commons formation and better enable stakeholders and policy makers to develop socially-valuable commons structures.

VI. CONCLUSION

While the sharing of data among scientists is necessary for the advancement of science and reflects a long-held norm of scientific practice, barriers and delays, legal, institutional, and interpersonal, have hindered the rapid and free sharing of scientific data. Stakeholder groups that are involved in the generation, funding, dissemination, and use of scientific data have policy objectives that are often at odds with one another and that frequently impact the design of scientific “commons.” Such objectives include the desire to use and access the findings of others, the desire to achieve recognition from one’s discoveries, the protection of commercial and financial interests, and the enablement of scientific advancement through broad availability of data.

The “latency analysis” that I propose in this Article is a tool that can be used to mediate among divergent stakeholder policy goals by defining two key variables that describe the temporal characteristics of information commons: the speed at which data enters the commons (knowledge latency) and the speed at which data within the commons becomes freely usable (rights latency). By placing latency analysis within larger theoretical frameworks for analyzing information commons, including the IAD framework championed by Ostrom and Hess, and the refined framework of Madison, Frischmann, and Strandburg, I provide a theoretical basis for the introduction and use of this methodology.

The three-part latency analysis methodology is useful not only in the analysis of existing commons structures, but in the design of new commons as well. In commons design, the introduction of latency variables to the negotiation over policy terms can effectively mediate among divergent stakeholder interests and objectives in a value-neutral manner, thus

to print and distribute the news reported in their papers. While the 1884 Bill failed, it laid the groundwork for the Supreme Court’s decision in *International News Service v. Associated Press*, 248 U.S. 215 (1918), recognizing a non-copyright tort of “hot news” misappropriation and leading to numerous cases in which originators of factual compilations (sports scores, stock quotes, etc.) are afforded an initial “exclusivity” period in which to enjoy the fruits of their labor before otherwise unprotectable information enters the public domain. *See, e.g.*, Richard A. Posner, *Misappropriation: A Dirge*, 40 HOUSTON L. REV. 621, 629–34 (2003).

facilitating compromise and commons formation. The case studies described in this Article—the genome commons and scientific publishing—exemplify the use of latency variables to achieve resolution of otherwise difficult value-based positions. Through the use of timing mechanisms, competing stakeholder interests in publication priority versus scientific advancement (in the case of the genome commons), and broad information accessibility versus protection of financial interests (in the case of scientific publishing) have been successfully mediated. I propose that latency analysis and design techniques should be adopted more broadly and explicitly to other information commons, particularly in the sciences, and that the general application and recognition of this methodology can improve both the efficiency of commons formation and the ultimate utility of the resulting commons for all stakeholders.