

ARTICLE

RESOLVING THE OPEN SOURCE PARADOX IN BIOTECHNOLOGY: A PROPOSAL FOR A REVISED OPEN SOURCE POLICY FOR PUBLICLY FUNDED GENOMIC DATABASES

*Donna M. Gitter**

TABLE OF CONTENTS

I.	INTRODUCTION	1476
II.	THE INTERNATIONAL HAPMAP PROJECT: SCIENTIFIC BACKGROUND AND DATA ACCESS POLICY	1480
III.	ANALYSIS OF THE SHORTCOMINGS OF THE INTERNATIONAL HAPMAP PROJECT DATA ACCESS POLICY	1486
	A. <i>The HapMap Project's Open Source Data Access Policy Does Not Actually Preclude Patenting by Licensees Who Violate the Policy</i>	1486

* Assistant Professor of Legal and Ethical Studies, Fordham University Schools of Business; J.D., University of Pennsylvania Law School; B.A., Cornell University. E-mail: gitter@fordham.edu. I am grateful to Fordham University for awarding me a Faculty Fellowship and also a Summer Research Stipend that together permitted me to complete this work. I appreciate the generosity of several experts who lent their insights to this Article: Mr. Lee Bendekgey, Senior Vice President and General Counsel, Nuvelo, Inc.; Dr. Jean McEwen, Program Director, Ethical, Legal, and Social Implications Program, National Human Genome Research Institute, National Institutes of Health; and Dr. Tim Peakman, Executive Director, U.K. Biobank. I owe special thanks to Professor Michael Malinowski, Ernest R. and Iris M. Eldred Professor and Associate Director and Co-Founder, Program in Law, Science and Public Health, at the Paul M. Hebert Law Center at Louisiana State University, for his attention to and helpful comments regarding my work. I also thank Samuel Mok for his research assistance.

- B. *The HapMap Project’s Open Source Data Access Policy Does Not Bind Third Parties Who Obtain the Data Through Means Other Than the HapMap Website*1488
- C. *The HapMap Project’s Open Source Data Access Policy Lacks a Clear Enforcement Mechanism and Suitable Remedy*1489
- D. *The HapMap Consortium Might Be Unable to Enforce Its ClickWrap Data Access License Internationally*1491

IV. ANALYSIS OF AN ENHANCED OPEN SOURCE DATA ACCESS POLICY FOR PUBLICLY FUNDED GENOMIC DATABASES1491

- A. *Description of the Proposed Enhanced Open Source Licensing Scheme*1492
- B. *The Proposed Enhanced Open Source Licensing Scheme Has the Potential to Solve the Problems Arising from the Original HapMap Licensing Agreement*.....1495
- C. *The Proposed Enhanced Open Source Licensing Scheme Is Feasible*1498
 - 1. *The Proposed Open Source Licensing System Is Reconcilable with International Agreements Favoring Data Release into the Public Domain* ..1498
 - 2. *Little Harm to Scientific Progress Is Likely if the Enhanced Open Source Licensing Scheme Proposed Here Were to Place Genome Sequencing Centers in a Privileged Position*.....1502
 - 3. *There Is a Market for Nonexclusive Licenses for Publicly Funded Genetic Databases*1507
 - 4. *The HapMap Database Qualifies for Trade Secret Protection and the Secrecy of the Data Likely Can Be Maintained*.....1509
 - 5. *The Enhanced Open Source Access Model Proposed Here Will Attract the Sort of Small, Distributed Contributions That Advance the Goals of Scientific Research*.....1516

V. CONCLUSION.....1519

I. INTRODUCTION

Several current biotechnology research initiatives launched by government and nonprofit groups have sought to make their

data freely accessible to others, so as to stimulate innovation.¹ Many of these initiatives have adopted the “open source” model² that has achieved prominence in the computing industry.³ In brief, an open source software development project will make computer source code⁴ publicly available for licensees to use, modify, and redistribute, provided that these licensees make their enhancements available to others on the same terms, an approach known as “copyleft.”⁵

1. See Press Release, Nat'l Human Genome Research Inst., Nat'l Insts. of Health, International HapMap Consortium Widens Data Access (Dec. 10, 2004) [hereinafter NIH Press Release], available at <http://www.nih.gov/news/pr/dec2004/nhgri-10.htm> (noting that both the Human Genome Project and the International HapMap Consortium made most of their data freely available on the Internet to “researchers to use in their efforts to find genes”); News Advisory, Nat'l Human Genome Research Inst., Nat'l Insts. of Health, International Consortium Launches Genetic Variation Mapping Project (Oct. 2002) [hereinafter NIH News Advisory], available at <http://genome.gov/10005336> (describing the International HapMap Project as a “public-private effort”).

2. See CAMBIA, THE CAMBIA BIOS INITIATIVE: BIOLOGICAL INNOVATION FOR OPEN SOCIETY 6 (2004), available at <http://www.bios.net/daisy/bios/10/version/live/part/4/data> (noting that “[o]pen innovation is becoming a strikingly successful model in Open Source Software” and proposing to “produce high quality and relevant biological technologies . . . and secure these technologies in a new, protected, universally-accessible commons”); Ensembl Project (Oct. 2006), <http://oct2006.archive.ensembl.org/info/about/project.html> (describing Ensembl as a joint project between the European Bioinformatics Institute and the Wellcome Trust Sanger Institute “to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes,” and emphasizing its commitment to “[r]elease of data and analysis into the public domain immediately” and “[o]pen, collaborative software development” without “restrictions on access to, or use of, the data provided and the software used to analyse and present it”); see also Kenneth Neil Cukier, *Community Property: Open-Source Proponents Plant the Seeds of a New Patent Landscape*, ACUMEN J. LIFE SCI., Oct.–Nov. 2003, at 54, 57 (citing many biotechnology initiatives that have adopted an open source model, along with others that have simply released all their information into the public domain).

3. See Josh Lerner & Jean Tirole, *The Economics of Technology Sharing: Open Source and Beyond* 2 (Nat'l Bureau of Econ. Research, Working Paper No. 10956, 2004), available at <http://ssrn.com/abstract=629598> (describing computer software as “[t]he most prominent example of open source production”).

4. Source code is a computer program in its original form, written and readable by human beings. Janet Elizabeth Hope, Open Source Biotechnology 66 (Dec. 23, 2004) (unpublished Ph.D. thesis, Australian National University), available at <http://opensource.mit.edu/papers/hope.pdf>. Because computers can execute only instructions coded as a series of binary numbers (ones and zeroes), source code must be “translated by means of another program into binary form, known as machine or object code.” *Id.*

5. See *id.* at 68 (explaining the copyleft licensing scheme developed in the software community and describing it as “an ingenious twist on the conventional copyright licence”); see also Lerner & Tirole, *supra* note 3, at 2 (“In an open-source project, . . . a body of original material is made publicly available for others to use, under certain conditions. In many cases, anyone who makes use of the material must agree to make all enhancements to the original material available under these same conditions.”).

As noted by Professors Lerner and Tirole, the term “copyleft” arose to distinguish it from the term “copyright,” “because if copyright seeks to keep intellectual

Scholars have described the open source model as a promising one for biotechnology, at least for upstream data such as genetic sequence databases.⁶ Previously, the most notable publicly funded effort aimed at genome sequencing, the Human Genome Project (HGP),⁷ followed a traditional public domain model and made all of its data freely available on the Internet, declining to impose any restrictions whatsoever on its use.⁸ Many commentators feared that this public domain model would permit commercial users to diminish the utility and accessibility of the HGP's public domain data by making proprietary their improvements to that data,⁹ a process some term "parasitic patenting."¹⁰ Consequently, a subsequent, related project, the

property private, copyleft seeks to keep intellectual property free and available." *Id.* at 6. However, Ms. Hope notes that open source software licensors sometimes do charge fees for their software. Hope, *supra* note 4, at 67–68 & n.12 ("Free software' is a matter of liberty, not price."). For a detailed account of the development and functioning of the open source software movement, see *id.* at 65–93.

6. See, e.g., Arti K. Rai, "Open and Collaborative" Research: A New Model for Biomedicine, in INTELLECTUAL PROPERTY RIGHTS IN FRONTIER INDUSTRIES: BIOTECHNOLOGY AND SOFTWARE 131, 152 (Robert Hahn ed., 2005) ("[W]hen the data in question are upstream, a significant case can be made in favor of publicly funded, publicly available databases that can be improved on collaboratively."); see also David W. Opperbeck, *The Penguin's Genome, or Coase and Open Source Biotechnology*, 18 HARV. J.L. & TECH. 167, 184 (2004) ("It seems clear that . . . gene sequence information must be open if biotechnological development involving genetic engineering is to be open source.").

7. Begun in 1990, the Human Genome Project (HGP) was a successful "13-year effort coordinated by the U.S. Department of Energy and the National Institutes of Health" to, among other things, "identify all the approximately 20,000–25,000 genes in human DNA, determine the sequences of the 3 billion chemical base pairs that make up human DNA, [and] store this information in databases." Human Genome Project, U.S. Dep't of Energy Office of Sci., About the Human Genome Project, http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml (last visited Jan. 10, 2007).

8. See Human Genome Project, U.S. Dep't of Energy Office of Sci., Genetics and Patenting, http://www.ornl.gov/sci/techresources/Human_Genome/elsi/patents.shtml (last visited Jan. 10, 2007) ("All genome sequence generated by the [HGP] has been deposited into GenBank, a public database freely accessible by anyone with a connection to the Internet."); see also Amy Kapczynski et al., *Addressing Global Health Inequities: An Open Licensing Approach for University Innovations*, 20 BERKELEY TECH. L.J. 1031, 1072 (2005) (explaining that the HGP "adhered to the classic public domain model and dedicated its outputs to the public without further restriction").

As noted by Professors Lerner and Tirole, in the software context, open source code differs from code placed in the public domain in that, for public domain code, "no restrictions are placed on subsequent users of the source code: those who add material in the public domain do not commit to put the new product in the public domain." Lerner & Tirole, *supra* note 3, at 6.

9. See Rai, *supra* note 6, at 142–43 ("The HapMap project releases individual genotype data as soon as it is identified. Before haplotype information has been assembled, it may be possible for those who access the data to take these data, combine them with their own genotype data, and generate enough information to file patent applications on haplotypes of interest.").

10. See NIH Press Release, *supra* note 1 (describing "parasitic patenting" as the

International HapMap Project (“HapMap Project”),¹¹ adopted an open source data access policy, at least at the outset.¹² According to a HapMap official, the HapMap Project’s initial open source approach is likely to serve as a model for future publicly funded genome sequencing efforts coordinated by the National Human Genome Research Institute of the National Institutes of Health.¹³

Some commentators contend that the open source data access policy adopted by the International HapMap Consortium

process used by “outside groups . . . [to] combine some of the HapMap data with their own data to generate patentable inventions,” and noting that “[s]uch patents could potentially be used to exclude other researchers from being able to freely use the HapMap data”). For a detailed description of the process of parasitic patenting, see Rebecca S. Eisenberg, *Genomics in the Public Domain: Strategy and Policy*, 1 NATURE REV. GENETICS 70, 73 (2000) (describing how private firms can use public data to enhance the value of their private data by incorporating the public data into inventions for which these private firms then file patent applications and declaring that “prompt disclosure in the public domain can be treacherous if your ultimate goal is to keep information freely available”).

11. See NIH News Advisory, *supra* note 1 (“Where the [HGP] provided the foundation on which researchers are making dramatic genetic discoveries, the HapMap will begin to make the results of genomic research applicable to individuals.”); Int’l HapMap Project, About the HapMap, <http://www.hapmap.org/thehapmap.html.en> (last visited Jan. 10, 2007) (“The International HapMap Project is a multi-country effort to identify and catalog genetic similarities and differences in human beings. Using the information in the HapMap, researchers will be able to find genes that affect health, disease, and individual responses to medications and environmental factors.”). The HapMap Project received funding from both public and private sources. See Int’l HapMap Project, Participating Groups, <http://www.hapmap.org/groups.html> (last visited Jan. 10, 2007) [hereinafter HapMap Participating Groups] (listing funding agencies).

12. See Int’l HapMap Project, Data Access Policy for the International HapMap Project, (on file with the Houston Law Review) (setting forth the HapMap’s data access policy and describing it as one “where some data are released quickly without restriction and some data are released with restrictions for a limited period of time”); see also Wash. Univ. in St. Louis Sch. of Med., Terms and Conditions for Data Access, SNP Research Facility Public Access License, <http://snp.wustl.edu/data/terms-and-conditions.html> (last visited Jan. 10, 2007) (including much of the HapMap Data Access Policy verbatim as its own policy). As explained by Dr. Jean McEwen, “[t]he concern was that investigators might be able to take the HapMap data deposited in the [public domain and combine it with a small amount of their own data and then try to file for patents on this very basic data – thus potentially impeding downstream development.” E-mail from Dr. Jean McEwen, Program Dir., Ethical, Legal, & Soc. Implications Program, Nat’l Human Genome Research Inst., Nat’l Insts. Of Health, to Donna M. Gitter, Assistant Professor of Legal & Ethical Studies, Fordham Univ. Sch. of Bus. (June 9, 2006, 11:38 EST) (on file with the Houston Law Review).

Eventually the International HapMap Consortium found that it was able to release all of its data into the public domain without restriction. See NIH Press Release, *supra* note 1 (“The International HapMap Consortium today announced that it is ending computer-based ‘click wrap’ license restrictions on data generated by its effort to create a map of human genetic variation. As a result, all of the consortium’s data are now completely available to the public . . .”). For a discussion of the reasons that the HapMap Consortium discontinued its initial open source policy, see *infra* notes 36–40 and accompanying text.

13. Telephone Interview with Dr. Jean McEwen, Program Dir., Ethical, Legal, & Soc. Implications Program, Nat’l Human Genome Research Inst., Nat’l Insts. Of Health (Nov. 21, 2005).

fails to protect adequately against the dangers of parasitic patenting.¹⁴ This Article analyzes the shortcomings of the HapMap Project's open source policy and proposes an enhanced model that preserves some features of the open source approach while providing additional protection against parasitic patenting. In order to lay the groundwork for this analysis, Part II provides a brief background of the aims and data access policy of the HapMap Project. Part III describes the significant shortcomings of the HapMap data access policy. Part IV then proposes an alternative data access policy that preserves some features of the open source approach while also providing enhanced protection against parasitic patenting. This Part both explains how the data access policy proposed here can alleviate the shortcomings of the initial HapMap policy and also demonstrates the feasibility of the plan. The alternative data access policy proposed here will help to achieve the aims of the National Human Genome Research Institute of the National Institutes of Health, which has emphasized its commitment to implementing data access policies that reduce the incidence of parasitic patenting, with the ultimate goal of advancing public health by making human genomic databases widely available to researchers worldwide.¹⁵

II. THE INTERNATIONAL HAPMAP PROJECT: SCIENTIFIC BACKGROUND AND DATA ACCESS POLICY

In October 2002, an international consortium launched the HapMap Project, "an approximately \$100 million public-private effort to create the next generation map of the human genome" with the purpose of "speeding the discovery of genes related to common illnesses such as asthma, cancer, diabetes and heart disease."¹⁶ Publishing results in October 2005, the HapMap Project succeeded in creating "a public database of common variation in the human genome."¹⁷ Using this database that catalogs genetic differences among individuals, future

14. See *supra* note 9 and accompanying text.

15. Telephone Interview with Dr. Jean McEwen, *supra* note 13.

16. NIH News Advisory, *supra* note 1. For a list of the various researchers and funding sources involved in the HapMap Project, see HapMap Participating Groups, *supra* note 11.

17. Int'l HapMap Consortium, *A Haplotype Map of the Human Genome*, 437 NATURE 1299, 1299 (2005) (reporting "a public database of common variation in the human genome" and showing how it "can guide the design and analysis of genetic association studies, shed light on structural variation and recombination, and identify loci that may have been subject to natural selection during human evolution").

researchers will be able to determine which genetic combinations contribute to many common multifactorial genetic disorders.¹⁸

In order to create the HapMap, researchers involved in the consortium effort first extracted DNA from blood samples gathered from voluntary donors in four countries: Nigeria, Japan, China, and the United States.¹⁹ These donors received adequate information so that they could give informed consent to the use of their samples.²⁰ In addition, researchers collected “[n]o medical or personal identifying information,” though the samples were “identified by the population from which they were collected.”²¹

Once collected by consortium researchers, the blood samples were “stored at the Coriell Institute for Medical Research in Camden, N.J., a nonprofit biomedical research center that specializes in storing living cells and making them available to scientists for further study.”²² Consortium researchers then analyzed the data over a three-year period to create the HapMap, culminating with the 2005 publication of their results in the journal *Nature*.²³

The National Human Genome Research Institute has described the HapMap’s creation and its value to subsequent researchers as follows:

The DNA sequence of any two people is 99.9 percent identical. The variations, however, may greatly affect an individual’s disease risk. Sites in the DNA sequence where

18. See NIH News Advisory, *supra* note 1 (noting that “[b]y comparing genetic differences among individuals,” the HapMap will allow researchers “to study the genetic risk factors underlying a wide range of diseases and conditions” that are caused both by genetic variants and other “[e]nvironmental and other non-genetic factors”).

19. Int’l HapMap Project, HapMap Sample Populations, <http://www.hapmap.org/hapmappopulations.html.en> (last visited Jan. 10, 2007) (noting that the blood samples came from the Yoruba people of Ibadan, Nigeria; individuals from Tokyo, Japan; individuals from Beijing, China; and U.S. residents with northern and western European ancestry, and that these blood samples would be “converted into cell lines . . . to make DNA”); see also Int’l HapMap Project, Ethical Concerns, <http://www.hapmap.org/ethicalconcerns.html.en> (last visited Jan. 10, 2007) (describing the consent process employed by the HapMap Project).

20. See Int’l HapMap Project, Consent Forms, <http://www.hapmap.org/consent.html.en> (last visited Jan. 10, 2007) (describing the process by which an informed consent template was modified “to make the consent form used in each location culturally appropriate”).

21. NIH News Advisory, *supra* note 1.

22. *Id.* As explained by Professor Clayton, these samples “will be made available to any investigator throughout the world who has an Institutional Review Board (‘IRB’)-approved protocol and who is approved by the Coriell IRB.” Ellen Wright Clayton, *Implications for Existing Law/Regulations*, 66 LA. L. REV. 125, 128 (2005). In addition, community advisory groups among the various populations will be able to remove samples from the repository if they object to the research conducted. *Id.*

23. See Int’l HapMap Consortium, *supra* note 17.

individuals differ at a single DNA base are called single nucleotide polymorphisms (SNPs). Sets of nearby SNPs on the same chromosome are inherited in blocks. This pattern of SNPs on a block is a haplotype. Blocks may contain a large number of SNPs, but a few SNPs are enough to uniquely identify the haplotypes in a block. The HapMap is a map of these haplotype blocks and the specific SNPs that identify the haplotypes are called tag SNPs.

The HapMap should be valuable by reducing the number of SNPs required to examine the entire genome for association with a phenotype from the 10 million SNPs that exist to roughly 500,000 tag SNPs. This will make genome scan approaches to finding regions with genes that affect diseases much more efficient and comprehensive, since effort will not be wasted typing more SNPs than necessary and all regions of the genome can be included.

In addition to its use in studying genetic associations with disease, the HapMap should be a powerful resource for studying the genetic factors contributing to variation in response to environmental factors, in susceptibility to infection, and in the effectiveness of and adverse responses to drugs and vaccines.²⁴

In terms of making the HapMap data available to users, the International HapMap Project emphasized its commitment to “rapid and complete data release, and to ensuring that Project data remain freely available in the public domain, at no cost to users.”²⁵ During the period from November 2003 to January 2006, the HapMap Consortium performed at least twenty separate data releases, which are available to anyone with an Internet connection.²⁶

In order to make certain that the HapMap data would remain accessible to all users, the HapMap Project explained at the outset that it “had to adopt a Data Release Policy where some data are released quickly without restriction and some data are released with restrictions for a limited period of time.”²⁷ Thus, the HapMap Project implemented “a free, non-exclusive, non-royalty-bearing licensing agreement to obtain

24. Nat'l Human Genome Research Inst., Nat'l Insts. of Health, International HapMap Project Overview, <http://www.genome.gov/10001688> (last visited Jan. 10, 2007).

25. Int'l HapMap Project, *supra* note 12.

26. Int'l HapMap Project, Old News, http://www.hapmap.org/old_news.html.en (last visited Jan. 10, 2007) (providing links to all twenty previous HapMap public releases).

27. Int'l HapMap Project, *supra* note 12.

access to certain types of data the project had collected on individuals' DNA sequences, specifically the genotypes."²⁸ Basically, the only condition of the license was that "users agreed not to prevent others from using the individual genotype data and to share data only with those who had also agreed to this condition."²⁹

The HapMap Consortium explained that its data access policy was formulated to avoid the filing of intellectual property claims that would impede other users' access to the data. As explained on the HapMap Project website that permits registration for access to the HapMap Project Genotype Database via the click of a mouse:

In formulating its data release policy, the HapMap Project had to address the potential problem of other parties filing, and being awarded, patents claiming certain of the data that the Project will produce. This concern arises because the genotype data produced during the early stages of the Project will not be sufficiently dense to allow derivation of haplotype information. During this time, however, it would be possible for others to combine the public HapMap Project's genotype data with their own, to construct haplotypes, to file for patents on those derived haplotypes, and in doing so potentially restrict others from using those haplotypes and underlying data. Therefore, the current licensing strategy has been adopted as a safeguard to prevent any such third-party patents that might otherwise use Project data being obtained or enforced in a way that would prevent others from using data generated by the Project, while at the same time allowing anybody who agrees to not use the data this way to have access to the data soon after they are generated.³⁰

The terms of the International HapMap Project Public Access License, as they appear on the HapMap Project website, are straightforward.³¹ The license provides that users may "access and conduct queries of the Genotype Database and copy, extract, distribute or otherwise use copies of the whole or any part of the Genotype Database's data as [they] receive it,

28. NIH Press Release, *supra* note 1.

29. *Id.*

30. Int'l HapMap Project, *supra* note 12.

31. See Int'l HapMap Project, International HapMap Project Public Access License (Aug. 2003) (on file with Houston Law Review) (setting forth the terms of the International HapMap Project Public Access License).

in any medium and for all (including for commercial purposes,” provided that they do not “restrict the access to, or the use which may be made by others of, the Genotype Database or the data that it contains.”³² Specifically, users are proscribed from filing patent applications that purport to patent

any single nucleotide polymorphism (“SNP”), genotype or haplotype data obtained from the Genotype Database or any SNP, haplotype or haplotype block based on data obtained from the Genotype Database; and [likewise from filing patent applications that] . . . contain claims to particular uses of any SNP, genotype or haplotype data obtained from the Genotype Database or any SNP, haplotype or haplotype block based on data obtained from, the Genotype Database, unless such claims do not restrict . . . the ability of others to use at no cost the Genotype Database³³

In addition, users may “disclose data obtained as a result of . . . access to and use of the Genotype Database only to other parties who have first confirmed . . . in writing that they too are licensees under the terms of the International HapMap Project Public Access License and so are bound by equivalent terms and conditions.”³⁴ Users are reminded that although they are not required to accept the license, by clicking on the “Accept” button at the end of the license, a user indicates assent with the license terms.³⁵

The HapMap Consortium stressed that its license was not intended to block the ability of users to file for intellectual property protection on specific haplotypes for which they have identified associated phenotypes, such as disease susceptibility, drug responsiveness, or other biological utility, as long as public access to, and use of, the data produced by the HapMap Project is preserved.³⁶

In other words, researchers could file patents once they were able to link a particular haplotype with an associated physical characteristic. The Consortium also announced its intention to

32. *Id.* § 2–2(a).

33. *Id.* § 2(b).

34. *Id.* § 2(c).

35. *Id.* § 4. Likewise, a party who does not accept the license lacks “permission to access, use, modify, distribute or otherwise use the Genotype Database or the data contained in it.” *Id.*

36. Int’l HapMap Project, *supra* note 12.

discontinue its licensing policy as soon as it was practicable, likely at the end of 2005.³⁷

Indeed, on December 10, 2004, the International HapMap Consortium announced that it would end its licensing policy, with the result that all the consortium's data would from that time forth be available to the public without restriction.³⁸ This change in policy resulted from various scientific advances in the field which "led the consortium to conclude that the patterns of human genetic variation can readily be determined clearly enough from the primary genotype data to constitute prior art" and therefore "derivation of haplotypes and 'haplotype tag SNPs' from HapMap data should be considered obvious and thus not patentable."³⁹ As a result, the consortium deemed its public access license no longer necessary and opted to post all of the consortium's monthly releases of data and to distribute this data to other public databases.⁴⁰ Within one year of changing its licensing policy, the HapMap Project published its full results in the journal *Nature*.⁴¹

A representative of the HapMap Project has indicated the Consortium's satisfaction with the HapMap Public Access License,⁴² which, as Professor Rai has noted, "is self-consciously modeled on the 'copyleft' system of open source software licensing."⁴³ Thus, the U.S. National Human Genome Research

37. *See id.* ("Once the HapMap Project has ensured that most of the data it has generated . . . have been placed in the public domain as haplotypes are derived, the need for the licensing approach to data release will disappear and the click-wrap step for access will be abandoned. The project anticipates that it will reach this goal around the end of . . . 2005.")

38. NIH Press Release, *supra* note 1.

39. *Id.*

40. *See id.*

41. Int'l HapMap Consortium, *supra* note 17, at 1299.

42. Telephone Interview with Dr. Jean McEwen, *supra* note 13. According to Dr. McEwen, in the case of the HapMap Project, parasitic patenting did not prove to be as much of a practical problem as expected. *Id.* She reported that the Project did not receive reports of any violations of its license or of parasitic patenting, and explained that the Project had appointed an intellectual property committee to monitor this issue. *Id.*

One reason for the absence of parasitic patenting claims may be the fact that in January 2001, the U.S. Patent and Trademark Office (PTO) implemented a heightened utility standard for patent examiners reviewing DNA sequence patents. *See* Utility Examination Guidelines, 66 Fed. Reg. 1092, 1093, 1098 (Jan. 5, 2001) (requiring patent application for purified and isolated gene to demonstrate a "specific, substantial, and credible" utility). The revised PTO guidelines render it less likely that a patent applicant would obtain a patent on a haplotype or other gene-based invention unless the applicant had also identified the associated phenotypes. *See supra* text accompanying note 36.

43. Rai, *supra* note 6, at 132 & n.6 (noting that the International HapMap Project Public Access License acknowledges the model of the GNU General Public License); *see also* Int'l HapMap Project, *supra* note 31 (posting at the beginning of the online license "[f]ull acknowledgements to the GNU General Public License Copyright © 1989, 1991

Institute, which coordinated the HapMap Project, is likely to implement a similar open source license for future publicly funded genomic research efforts. This Article posits that the open source license used by the HapMap Project suffers from several significant shortcomings that render it inadequate to prevent the dangers of parasitic patenting, and proposes an alternative for future publicly funded genomic databases. The proposed alternative also relies on an open source framework but provides enhanced protection in order to keep such data widely accessible to all users.

III. ANALYSIS OF THE SHORTCOMINGS OF THE INTERNATIONAL HAPMAP PROJECT DATA ACCESS POLICY

A. *The HapMap Project's Open Source Data Access Policy Does Not Actually Preclude Patenting by Licensees Who Violate the Policy*

As a threshold matter, some commentators have expressed concern regarding the inadequacy of the HapMap Project's data access license to prevent prohibited patenting activity.⁴⁴ This stems from the fact that the HapMap license is no more than a contractual agreement between two parties.

Analogizing the HapMap Public Access License to the General Public License (GPL),⁴⁵ courts would likely consider the GPL a "clickwrap" agreement.⁴⁶ One federal district court has defined a clickwrap as an agreement that presents the end-user "with a message on his or her computer screen, requiring that the user manifest his or her assent to the terms of the license agreement by

Free Software Foundation, Inc.").

As explained by Ms. Hope, the GNU General Public License (GPL), drafted by the Free Software Foundation (FSF), is the "archetypal copyleft licence." Hope, *supra* note 4, at 67–68. The FSF is a nonprofit group that seeks to "preserve, protect and promote the freedom to use, study, copy, modify, and redistribute computer software, and to defend the rights of all free software users." Free Software Found., About Us, <http://www.fsf.org/about> (last visited Jan. 10, 2007). For more information on the FSF and the GNU General Public License, see Free Software Found., <http://www.fsf.org> (last visited Jan. 10, 2007).

44. See, e.g., Opderbeck, *supra* note 6, at 198–200 (expressing skepticism as to the applicability of an open source model to biotechnology, and to the HapMap in particular).

45. See *supra* note 43 and accompanying text.

46. See Jason B. Wacha, *Taking the Case: Is the GPL Enforceable?*, 21 SANTA CLARA COMPUTER & HIGH TECH. L.J. 451, 488 (2005) ("Based on the way that the GPL is usually delivered to licensees, it is common to analyze its enforceability as a clickwrap or shrinkwrap agreement."); see also David McGowan, *Legal Implications of Open-Source Software*, 2001 U. ILL. L. REV. 241, 289 (likening the GPL to a shrinkwrap agreement, which is similar to a clickwrap agreement).

clicking on an icon. The product cannot be obtained or used unless and until the icon is clicked.”⁴⁷ Commentators analyze clickwrap agreements as contracts,⁴⁸ and note that courts generally enforce these contracts so long as they fulfill certain requirements.⁴⁹

More specifically, courts generally enforce clickwrap agreements provided the licensee “receive[s] notice of the license terms before buying or using the software”; has “the ability to return the software without using it, or to not download it, if he does not agree with the terms”; and “take[s] some definitive act to accept the terms.”⁵⁰ The International HapMap Project Public Access License fulfills all of these conditions, as the license was displayed before a potential user could access the data; a potential licensee could decline to download the data after reading the policy; and the license required affirmative assent, providing “Accept” and “Decline” buttons and stating that by clicking the “Accept” button, the licensee “certified] that [he] ha[d] read and accept[ed] the terms and conditions above.”⁵¹

Notwithstanding the enforceability of the HapMap License under contract law, however, a court’s finding of breach of contract by a licensee who misused information in the HapMap database would not automatically invalidate any patent obtained by that licensee. As stated by Professor Opderbeck:

Nothing in the Patent Act would suggest that a patent could be invalidated because some of the underlying data

47. *Specht v. Netscape Commc’ns. Corp.*, 150 F. Supp. 2d 585, 593–94 (S.D.N.Y. 2001) (footnote omitted) (describing how clickwrap licensing works), *aff’d*, 306 F.3d 17 (2d Cir. 2002).

48. *See* Wacha, *supra* note 46, at 482 (noting that the obligation imposed by the GPL on a licensee to disclose source code “could only be enforceable under a contract theory, as opposed to a license theory”); *see also* Kevin W. Grierson, Annotation, *Enforceability of “Clickwrap” or “Shrinkwrap” Agreements Common in Computer Software, Hardware, and Internet Transactions*, 106 A.L.R.5th 309, 318–19 (2003) (“Most courts have analyzed such agreements under the principles of contract formation found in the Uniform Commercial Code”); McGowan, *supra* note 46, at 289–302 (analyzing shrinkwrap licenses under a contract theory).

49. *See* Wacha, *supra* note 46, at 488 (“It is relatively certain under U.S. law that shrinkwraps and clickwraps are legally enforceable as long as they meet certain requirements.”); Jane K. Winn, *Contracting Spyware by Contract*, 20 BERKELEY TECH. L.J. 1345, 1352–53 (2005) (citing “the strong trend in recent cases favoring the enforcement of clickwrap agreements in the absence of a conflict between contract terms and fundamental public policy of the forum, or evidence of misconduct so egregious that it might rise to the level of unconscionable” and stating that “[i]n all, more than a dozen cases have been decided upholding the enforceability of contracts formed using click-through interfaces”).

50. Wacha, *supra* note 46, at 488; *see also* Grierson, *supra* note 48, at 318–19 (indicating that courts commonly cite failure to satisfy the requisite conditions among their reasons for rejecting clickwrap or shrinkwrap licenses).

51. *See* Int’l HapMap Project, *supra* note 31.

was derived from a database in violation of the database's terms of use. Thus, it is unlikely that the HapMap license provides the HapMap Consortium any meaningful remedy once a patent has been filed.⁵²

Notwithstanding the fact that the HapMap Consortium cannot prevent a licensee who misuses its database from obtaining a patent, it is important to note that a licensee who misuses the HapMap data by engaging in parasitic patenting, for example, might find itself unable to pursue successful patent infringement lawsuits against others based upon the data it had used improperly in the first place. The alleged infringer could simply raise the defense that it had used data from the public domain, and therefore did not infringe the HapMap licensee's patent. Thus, the HapMap data access policy could in a sense deter misuse, since those who misappropriate the information cannot rest secure in patent rights obtained in this way. The drawback of this enforcement strategy, however, is that it places the HapMap Consortium in a fairly passive position, as it requires patenting activity by a third party and also a successful patent infringement defense by that third party.

B. The HapMap Project's Open Source Data Access Policy Does Not Bind Third Parties Who Obtain the Data Through Means Other Than the HapMap Website

Another shortcoming of the HapMap License is that it does not bind third parties who obtain and use the HapMap data

52. Opderbeck, *supra* note 6, at 199. "Four major substantive defenses can preclude enforcement of a patent against otherwise infringing conduct," including patent invalidity, fraudulent procurement or inequitable conduct, misuse or violation of the antitrust laws, and delay in filing suit resulting in laches or estoppel. 6 DONALD S. CHISUM, CHISUM ON PATENTS § 19.01 (2005). Research did not reveal any litigation on the issue, but none of these defenses would seem to apply to defeat the patent of a licensee who has obtained his patent by breaching a clickwrap agreement concerning use of a database. *See generally id.* In this respect a genetic database differs significantly from open source software because the latter remains protected by copyright even in the absence of an open source agreement. *See* 17 U.S.C. § 117 (2000) (recognizing copyright protection for computer programs); McGowan, *supra* note 46, at 289 ("If the GPL is ineffective, the copyright still persists."). In contrast, U.S. law does not provide copyright protection for genetic databases. *See* Edward J. Baba, Note, *From Conflict to Confluence: Protection of Databases Containing Genetic Information*, 30 SYRACUSE J. INT'L L. & COM. 121, 137 (2003) (explaining that, after the U.S. Supreme Court decision in *Feist Publications, Inc. v. Rural Telephone Services Co.*, 499 U.S. 340 (1991), courts generally hold that databases, such as genetic databases, do not merit copyright protection because they lack the requisite creativity and originality); *see also* Opderbeck, *supra* note 6, at 199 (noting that the HapMap data "in itself is not protected by copyright"); Rai, *supra* note 6, at 143 (noting that the HapMap Consortium cannot "rely on an assertion of copyright in the underlying data").

without downloading it from the HapMap website⁵³ and who therefore are not in privity of contract with the HapMap Consortium.⁵⁴ The drafters of the HapMap License attempt to avoid this problem by including a clause that permits a licensee to disclose HapMap data “only to other parties who have first confirmed to you in writing that they too are licensees under the terms of the International HapMap Project Public Access License and so are bound by equivalent terms and conditions” imposed by the HapMap License,⁵⁵ but this clause is equally ineffective against third parties, as by its terms it can bind only the original licensee.

C. The HapMap Project’s Open Source Data Access Policy Lacks a Clear Enforcement Mechanism and Suitable Remedy

Some scholars have also expressed doubt as to whether government or nonprofit bodies that create genetic databases will enforce their data access policies by initiating litigation, if necessary, against a party who breaches their license agreement.⁵⁶ In the case of the HapMap Project, the HapMap Consortium would have to bring suit in order to enforce its

53. For example, some researchers may obtain HapMap data from other researchers who initially did download the HapMap data from the HapMap website.

54. See Richard A. Mann & Barry S. Roberts, *CyberLaw: A Brave New World*, 106 DICK. L. REV. 305, 330 n.186 (2001) (stating “the enforceability of bilateral cyberspace contracts may be suspect for a lack of privity” (citing Robert P. Merges, *The End of Friction? Property Rights and Contract in the “Newtonian” World of On-Line Commerce*, 12 BERKELEY TECH L.J. 115, 119–20 (1997))); Opderbeck, *supra* note 6, at 198 (noting that “[e]ven in the context of open source software, it is unclear whether a GPL or Creative Commons type of license would be enforceable against downstream parties who are not in privity with the originator of the underlying work”); Rai, *supra* note 6, at 143 n.58 (“Because there is no underlying copyright, those who manage to access the data without having agreed to the license are not subject to any legal prohibition against patenting.”).

Professor Opderbeck notes that open source licenses for *software*, which is copyrightable, are therefore very distinct from open source licenses for *genetic databases*, which do not receive copyright protection. He explains that “[b]ecause of the nature of derivative works, which imply separate ownership rights in the underlying work and the ‘new’ portions of the derivative work, it is easy to imagine enforceable license terms that regulate distribution of derivative works and run with the copyright.” Opderbeck, *supra* note 6, at 199.

55. Int’l HapMap Project, *supra* note 31.

56. Cf. Symposium, *Open Source Genomics*, 8 B.U. J. SCI. & TECH. L. 254, 268 (2002) (quoting Professor Dan Burk, University of Minnesota Law School, as expressing doubt that, had the National Institutes of Health (NIH) succeeded in 1991–1992 in its efforts to patent thousands of human gene fragments of unknown utility, it would actually have pursued litigation to enforce such patents). For more information on the controversy surrounding the NIH’s pursuit of these patents in the past, see Donna M. Gitter, *Led Astray by the Moral Compass: Incorporating Morality into European Union Biotechnology Patent Law*, 19 BERKELEY J. INT’L L. 1, 11 n.78 (2001).

contract.⁵⁷ This would create a significant financial and administrative strain upon the nonprofit research group, which must focus its efforts on pursuing research as opposed to enforcing its data access policy.

Another enforcement problem with the HapMap License arises in terms of the measure of damages. As indicated by Professor Opderbeck, "Since the HapMap project is a nonprofit venture, any damages from such a contractual breach are likely to be highly speculative."⁵⁸ Even an equitable remedy such as an injunction preventing the licensee from selling any commercial product that it developed through use of the HapMap data would prove insufficient because the HapMap Consortium certainly does not seek to impede commercial development based upon its data. Indeed, the HapMap seeks to encourage commercialization, so long as licensees do not block others' use of the HapMap data.⁵⁹

As discussed above, however, one possible deterrent against parasitic patenting might be the fact that a HapMap licensee who misused the HapMap data in this way could not effectively defend its patent in court against alleged infringers.⁶⁰ Thus, the difficulties of enforcement by the HapMap Consortium and the lack of an effective remedy would prove less critical, as enforcement would essentially be carried out by third parties whose defense against infringement lawsuits would ultimately deprive licensees who misused the data from enjoying their ill-gotten gains. As pointed out above, however, this enforcement strategy places the HapMap Consortium in a fairly passive position, as it requires patenting activity by a third party.⁶¹

57. The situation in the open source software community is different because software created under the GPL enjoys copyright protection. *See supra* note 52. Thus, a party who holds a copyright in a work licensed under the GPL is "the only party who can seek a remedy for violation of the GPL as applied to that work." Brian W. Carver, Note, *Share and Share Alike: Understanding and Enforcing Open Source and Free Software Licenses*, 20 BERKELEY TECH. L.J. 443, 465 (2005). Because the FSF does indeed "hold the copyrights to many widely-used free software programs, particularly those within its GNU project," the FSF "does handle enforcement, primarily through its pro bono general counsel." *Id.*

58. Opderbeck, *supra* note 6, at 199.

59. *See* Int'l HapMap Project, Data Release Policy, <http://www.hapmap.org/datareleasepolicy.html> (last visited Jan. 10, 2007) ("[I]f a specific utility can be demonstrated for a SNP or haplotype, any group, whether associated with the Project or not, should be able to apply for a patent, as long as this action does not prevent others from obtaining access to data from the Project.").

60. *See supra* Part III.A.

61. *See id.*

D. The HapMap Consortium Might Be Unable to Enforce Its ClickWrap Data Access License Internationally

The HapMap database is accessible to any user equipped with a computer and Internet connection, no matter where that user is located. If the user happens to be located in a nation that does not enforce clickwrap licenses, then that user might not face legal liability for violating the HapMap license.⁶² For this reason, along with the others listed in Part III, there is significant need for a data access policy for publicly funded genomic databases that will ensure the data's accessibility while simultaneously protecting against parasitic patenting. Part IV proposes a promising alternative.

IV. ANALYSIS OF AN ENHANCED OPEN SOURCE DATA ACCESS POLICY FOR PUBLICLY FUNDED GENOMIC DATABASES

The shortcomings described above affecting the HapMap open source data access policy are particularly significant in light of the fact that, increasingly, U.S. government funding agencies are devoting resources toward genomic sequencing projects that involve creating large, widely accessible databases, while at the same time wishing to avoid the problems engendered by parasitic patenting.⁶³ This Part proposes an enhanced open source

62. See, e.g., LUCIE M.C.R. GUIBAULT, COPYRIGHT LIMITATIONS AND CONTRACTS 205–07, 297 (2002) (noting the dearth, at the time of writing, of European court decisions concerning the validity of shrinkwrap licenses, much less clickwrap licenses, and further explaining that “the European courts have adopted a much more circumspect attitude” toward electronic or distance contracts as compared to U.S. courts); Jorge L. Contreras, Jr. & Kenneth H. Slade, *Click-Wrap Agreements: Background and Guidelines for Enforceability*, 2000 COMPUTER UND RECHT INT'L 104, 108 (suggesting that, as of the time of writing, some jurisdictions might not honor clickwrap agreements).

It should be noted that a German district court in Munich is the first court that has enforced the GPL. In 2004, that court held the GPL to be a valid and enforceable copyright license under German law. Landgericht Munchen I [LG] [trial court] May 19, 2004, *Welte v. Sitecom Deutschland GmbH*, No. 21 0 6123/04, available at http://www.jbb.de/urteil_lg_muenchen_gpl.pdf (official publication in German) and http://www.jbb.de/judgment_dc_munich_gpl.pdf (unofficial translation in English).

63. See Nat'l Human Genome Research Inst., Nat'l Insts. of Health, Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-Scale Sequencing and Other Community Resource Projects (Feb. 2003), <http://www.genome.gov/10506537>.

Large resource data sets are becoming an increasingly critical component of biomedical and biological research and, as such, will be more frequently produced specifically as community resources. NHGRI will encourage, as an integral component of the development of the new community resources it will support, planners and participants to devise appropriate approaches to implement the principle and achieve the advantages of rapid pre-publication data release. While addressing important considerations as data quality standards, data storage and dissemination modes, protection from parasitic intellectual property claims, and producer and user interests, the development of

licensing scheme to be implemented by organizations pursuing large-scale genomic research, analyzes the effectiveness of this alternative, and then examines whether such an approach would engender some unintended problems of its own.

A. *Description of the Proposed Enhanced Open Source Licensing Scheme*

In implementing a data access policy modeled on the open source approach, the HapMap Consortium aimed not only to facilitate researchers' access to haplotype data but also to preempt others from patenting this data, which is nonsubstitutable and therefore cannot be invented around.⁶⁴ The lack of enforceability of such a data access policy, however, threatens to impede the HapMap data access policy from achieving its desired ends.⁶⁵ Therefore, creators of future large-scale, publicly funded genomic databases ought to implement a nonexclusive, nonroyalty-bearing licensing policy for such data.

A creator of such a database essentially could protect its data as a trade secret,⁶⁶ offering reasonably priced subscription agreements for nonexclusive access to the database and requiring the user to execute a confidentiality agreement prohibiting the user from revealing the data.⁶⁷ Subscriptions should be available to any interested parties and offered at varying rates, with noncommercial researchers enjoying reduced fees in order to facilitate their access to this data.⁶⁸ Moreover, database creators

effective means to achieve the objectives of the community resource concept will maximize the benefit to the entire scientific community and to research.

Id.

64. See Hope, *supra* note 4, at 155 (describing the human genome as a "non-substitutable" standard).

65. See *supra* Part III.C for a detailed discussion regarding the lack of enforceability of the HapMap Project.

66. See *infra* Part IV.C.4 for a detailed discussion of trade secret law as applied to the HapMap Project.

67. See 1 ROGER M. MILGRIM, MILGRIM ON TRADE SECRETS § 4.01 (2006) ("The principal right of the owner of a trade secret is to grant access to the secret known to others subject to a contractual duty not to use or disclose it."). In addition to requiring confidentiality agreements for the protection of its information, which would be available electronically, the HapMap Project could also implement encryption measures.

68. See *infra* Part IV.C.5 for a full exploration of the definition of noncommercial researchers and a description of the pricing system envisaged here. While for-profit firms might be expected to protest reduced fees for noncommercial researchers, examples abound where for-profit firms themselves grant public sector, academic, and nonprofit researchers access to their databases at reduced rates. See Cristina Weschler, Note, *The Informal Experimental Use Exception: University Research After Madey v. Duke University*, 79 N.Y.U. L. REV. 1536, 1553-54 (2004) (citing several examples of this phenomenon). The idea behind offering reduced fees to noncommercial researchers is to maximize the public good arising from partially publicly funded databases such as the

should require a grantback, meaning that—in the case of the HapMap Project—any entity that derives haplotypes and haplotype tag SNPs⁶⁹ would grant back to the database creator and to all other users of the database nonexclusive freedom to use these haplotypes and haplotype tag SNPs as tools for their own research.⁷⁰

The grantback strategy proposed here for databases is similar to the one Incyte Genomics, Inc., a for-profit corporation, successfully pioneered in the early 1990s for its genetic database.⁷¹ Randall Scott, the president of Incyte, testified that it “was the first company to license its [genetic database] on a nonexclusive basis to pharmaceutical researchers.”⁷²

Incyte’s strategy both circumvented the problem of parasitic patenting and also increased the value of Incyte’s data.⁷³ As explained by Lee Bendekgey, a former Incyte executive, with respect to that company’s licensing strategy:

“[I]t occurred to the folks at Incyte and Pfizer late in negotiations [for Incyte’s first database product] that Incyte could be inadvertently starting an arms race among the pharmaceutical companies in the following way. Like any good pharma company that lives and dies by its IP portfolio, one thing Pfizer insisted on—like basically all others, in negotiations—was that if they discovered the full-length gene based on a partial gene in our database, and figured out its function, they could patent that: it would be theirs, not ours. That seems fair, but then it started to occur to everyone that it was quite likely, given everyone was starting with the same data, that people would identify a

HapMap Project. It is important to keep in mind, however, as commentators have rightly noted, that “entrepreneurs also pay taxes on their profits” and therefore promote the public good in this way. See 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, 427 & n.560 (2003) (arguing in support of wide dissemination of government-funded data).

69. See *supra* text accompanying note 24 (defining haplotypes and haplotype tag SNPs).

70. As noted by one commentator, “non-exclusive grant-backs of licensee improvements are well known in conventional biotechnology licensing.” Hope, *supra* note 4, at 130–31.

71. For a detailed description of Incyte’s nonexclusive licensing approach for access to its database, see Rebecca S. Eisenberg, *Intellectual Property at the Public-Private Divide: The Case of Large-Scale cDNA Sequencing*, 3 U. CHI. L. SCH. ROUNDTABLE 557, 568–69 (1996). See also Hope, *supra* note 4, at 159–60.

72. *Gene Patents and Other Genomic Inventions: Hearing Before the Subcomm. on Courts and Intellectual Property of the H. Comm. on the Judiciary*, 106th Cong. 47 (2000) (statement of Dr. Randal W. Scott, President and Chief Scientific Officer, Incyte Genomics).

73. See Hope, *supra* note 4, at 160.

lot of the same partial genes as being of interest, and then you would get this kind of race in which everyone was racing to complete the gene and figure out its function and get the IP on it and then start lobbing legal missiles at all the others. [T]here started to be concern [that] people would be contaminated and face the risk of IP infringement actions from other people who were using the same data. So what was proposed, and what Pfizer agreed to in the first agreement, and every big pharmaceutical company has agreed to since then, is in effect an open source-like grant-back. What those agreements say is that if any of our database users discovers and characterises a full-length gene using the information in our database, they grant back to Incyte and to all other users of our database non-exclusive freedom to operate . . . to use it as a target for your own drug discovery. This was actually before open source, but . . . it's actually the same concept: you have all these people working with the same starting point, and if they generate what can be thought of as an improvement—you know, added value—it gets granted back to everybody else. For Incyte itself, . . . it removes the impediment [to selling the database], which was the original benefit, and now there is the additional benefit that it is perceived as an additional source of value: they're not just getting access to our IP, they are also getting access to the IP of all these other pharmaceutical companies who have been working with this data for some significant period of time. So that is really the benefit to us, it increases the value of the product we are offering. So that story illustrates that they can be persuaded to share, but it's not easy—and the closer you get to the end product the harder the task of persuading them will be."⁷⁴

Incyte's strategy could apply to publicly funded genetic databases such as the HapMap database, which is sufficiently upstream that pharmaceutical firms would recognize the benefits of sharing it with one another.⁷⁵

74. *Id.* at 159–60 (quoting Lee Bendekgey) (alterations in original) (omissions in original).

75. Professor Rai cites several examples where, “in recent years, various pharmaceutical companies and other downstream developers have endeavored to defeat property rights in upstream information by paying to place this information (for example, information on gene fragments as well as information on SNPs) in the public domain.” Arti K. Rai, *Fostering Cumulative Innovation in the Biopharmaceutical Industry: The Role of Patents and Antitrust*, 16 BERKELEY TECH. L.J. 813, 832 (2001) (citing, among others, Merck's partnership with Washington University to place gene fragments in the public domain and the SNP Consortium's efforts to place an SNP map in the public domain).

According to Professor Malinowski, however, it is important to keep in mind that an open source model might prove an anathema to commercial biopharmaceutical

B. The Proposed Enhanced Open Source Licensing Scheme Has the Potential to Solve the Problems Arising from the Original HapMap Licensing Agreement

As discussed above, the initial HapMap data access policy, which was modeled upon the GPL in the computer industry, suffers from several disadvantages that render it inadequate to protect against parasitic patenting.⁷⁶ It is useful to examine whether the enhanced open source model proposed here will rectify these shortcomings.

First, the alternative proposed here has the potential to effectively impede patenting by those who would violate the initial HapMap licensing agreement. Of course, no contract can secure its own enforcement. The best one can hope is that the contract imposes significant consequences for its violation. To implement the enhanced HapMap licensing agreement proposed here, the HapMap Consortium would have had to impose a penalty upon a violator by denying further access to the database upon the Consortium's discovery of the breach. Because the HapMap project involved ongoing releases of data every month or two,⁷⁷ denying access would be a meaningful deterrent to keep users from violating the terms of their access, lest they be denied opportunities for future innovation. Denial of access would have been impossible under the initial HapMap data access policy, since the data was freely available to anyone with an Internet connection willing to click (if not abide by) the "Agree" button on the HapMap website. Under the model proposed here, although the HapMap Consortium would incur some costs associated with

firms, which have, as he puts it, an "addiction" to "strong IP in conjunction with investing." E-mail from Michael Malinowski, Ernest R. & Iris M. Eldred Professor, Assoc. Dir. & Co-Founder of the Program in Law, Sci. & Pub. Health, Paul M. Hebert Law Ctr. at La. State Univ., to Donna M. Gitter, Assistant Professor of Legal & Ethical Studies, Fordham Univ. Sch. of Bus. (July 28, 2006, 11:54 EST) (on file with the Houston Law Review). He notes that an open source model could therefore potentially chill, rather than encourage, pharmaceutical research and development, but Professor Malinowski observes: "It may prove that the HapMap enabling tech[nology] is so powerful that this does not happen, but it's always the law and policy fear . . ." *Id.*

This Article maintains a focus on open source style licensing for the HapMap Project in light of the fact that the HapMap Project has itself implemented a variant of the open source approach, an approach which this Article contends is flawed in several respects.

76. See *supra* Part III.

77. See Int'l HapMap Project, *supra* note 26 (announcing public data releases which have occurred at least every two months since November 11, 2003); cf. Hope, *supra* note 4, at 164 (quoting Stephen Maurer as stating: "In the case of biological databases, tiny updates are all that are being sold, and they come out very frequently. Companies would pay a big premium in order to get Genbank every twenty-four hours . . ." (omission in original) (internal quotation marks omitted)).

enforcing its license, these costs could be defrayed by the licensing fees it would charge. Thus, enforcement would not hinge upon a database user bringing a successful defense to infringement against another database user who had misused the HapMap data by engaging in parasitic patenting.⁷⁸

Second, the enhanced open source model proposed here would bind third parties who obtained the data through means other than the HapMap website. The HapMap Consortium and its licensees would execute a formal agreement acknowledging the licensee's right to use the Consortium's proprietary information on the condition that the licensee maintain the information's confidentiality.

Certainly, it is difficult to preserve trade secret status and there is always concern that the secret will become available to others. Under trade secret law, however, privity of contract is not necessarily required, such that "[t]hird parties not in privity with a trade secret owner who obtain the trade secret may be liable for trade secret misappropriation if they know or had reason to know that they obtained the trade secret from someone who had acquired or disclosed it unlawfully."⁷⁹ It is quite likely that, were international consortia such as the HapMap Project to adopt the model proposed here, the existence of trade secret protection would be sufficiently widely known, at least in the scientific community, that third parties could be held to have received constructive notice of its status as a trade secret.⁸⁰ This differs significantly from the initial HapMap licensing agreement, which requires privity of contract and would therefore be unenforceable against a third party who obtained the information from a licensee who violated the license agreement.⁸¹

Third, a database creator such as the HapMap Consortium could enforce its rights under this model and win a significant

78. See *supra* Part III.A.

79. Michael J. Hutter, *The Case for Adoption of a Uniform Trade Secrets Act in New York*, 10 ALB. L.J. SCI. & TECH. 1, 21 (1999). The lack of protection provided by trade secret law against an "honest" discoverer is not a drawback in the case of the HapMap Project, which sought to promote innovation as opposed to profit from its discoveries. See 2 MILGRIM, *supra* note 67, § 7.02[1][a].

80. See *Computer Assocs. Int'l, Inc. v. Altai, Inc.*, 982 F.2d 693, 718 (2d Cir. 1992) (holding that in a misappropriation action, the third party does not need to have actual notice that the information it received was acquired unlawfully and that "constructive notice is sufficient").

81. See Int'l HapMap Project, *supra* note 12 (setting forth the requirement that database users may share information only with other parties who have agreed to the access license and creating no enforceable right against third parties who did not agree to the license).

measure of damages if successful in litigation.⁸² Part of the licensing fees collected by the Consortium could be devoted toward legal fees, if necessary, to enforce its rights under the license. In terms of damages, “[t]he most commonly accepted measure of damages for trade secret misappropriation is the defendant’s profits,” deriving from “a perceived policy against unjust enrichment.”⁸³ Indeed, “even where the plaintiff has not shown that he has suffered any loss on account of a misappropriation, he has been entitled to recover the defendant’s profits.”⁸⁴ This remedy is more generous than that available under a breach of contract theory. In addition, injunctive relief is available in trade secret litigation,⁸⁵ and may even be awarded against third parties who received the secret information from a party that was subject to a duty not to disclose it.⁸⁶

Finally, the HapMap license proposed here could be enforced internationally.⁸⁷ Commentators have noted that, in the European Union, while there is no uniformity in terms of the legal protection of trade secrets,⁸⁸ “most countries provide some type of civil and criminal protection against the misappropriation of confidential commercial information, whether by breach of a contract or other wrongful conduct.”⁸⁹ According to these commentators, “trade secret law in the [European Union] can provide partial, but often very useful, protection for databases.”⁹⁰ Another scholar, while observing that trade secret protection in the United Kingdom and Australia is not as strong as in the United States,⁹¹ nonetheless concluded that “many legal systems

82. See generally Michael A. Rosenhouse, Annotation, *Proper Measure and Elements of Damages for Misappropriation of Trade Secret*, 11 A.L.R.4th 12, 20 (1982) (laying out potential damages for trade secret infringement).

83. *Id.*

84. *Id.*

85. See 4 MILGRIM, *supra* note 67, § 15.02[1][a].

86. See *id.* § 15.02[1][b].

87. See Jacqueline Lipton, *Protecting Valuable Commercial Information in the Digital Age: Law, Policy and Practice*, 6 J. TECH. L. & POL’Y 1, 29–30 (2001) (noting an increasing integration of international laws relating to trademark protection).

88. Mary Maureen Brown et al., *Database Protection in a Digital World*, 6 RICH. J.L. & TECH. 2, ¶ 77 & n.138 (1999), <http://law.richmond.edu/jolt/v6i1/conley.html> (“There is no substantive EU trade secrets law.”).

89. *Id.* ¶ 77 & n.139 (citing examples of trade secret protection under French and German law).

90. *Id.* ¶ 77.

91. See Lipton, *supra* note 87, at 9–15 (recognizing that the United States provides outright protection of trademarks through legislation rather than judicially created causes of action as found in the United Kingdom and Australia).

are now not particularly far apart in terms of what they will and will not protect in terms of valuable commercial information.⁹²

C. The Proposed Enhanced Open Source Licensing Scheme Is Feasible

Though the proposed enhanced open source licensing scheme described here has the potential to resolve some problems engendered by the current approach, it is essential to examine whether the proffered alternative raises some problems of its own. These possible problems include concerns: (1) that the proposed policy conflicts with international agreements favoring data release into the public domain; (2) that scientific progress would suffer if this proposal were to place genome sequencing centers in a privileged position; (3) that there may exist no viable market for nonexclusive licenses for such data; (4) that it might prove exceedingly difficult to maintain the trade secrecy of such databases; and (5) that such a fee-based access model might not attract the sort of small, distributed contributions for which an open source approach is ideal. This Article concludes that these concerns, while significant, do not provide sufficient reason to retain the current data access policy for publicly funded genomic databases.

1. The Proposed Open Source Licensing System Is Reconcilable with International Agreements Favoring Data Release into the Public Domain. One challenge facing the licensing plan proposed here is that such a scheme would violate the Bermuda Statement, an international agreement favoring release into the public domain of genetic databases achieved through public funding.⁹³ In February 1996, the Bermuda Statement garnered the unanimous consent of the

92. See *id.* at 29 (citing RAYMOND T. NIMMER, *THE LAW OF COMPUTER TECHNOLOGY* ¶ 3.02[1] (3d ed. 1997)). For example, Dr. Lipton notes that courts in the United Kingdom and Australia, which rely on “notions of contract law and on ideas of good faith and fiduciary relationships” rather than on a property theory of trade secrets, nonetheless have enforced actions of breach of confidence “against dishonest third parties who have shared no relationship of confidence with the plaintiff.” *Id.* at 9–10. *But cf.* Jacqueline Lipton, *Balancing Private Rights and Public Policies: Reconceptualizing Property in Databases*, 18 *BERKELEY TECH. L.J.* 773, 820 (2003) (“Because of trade secrecy law’s national and international divergence and . . . its inherent shortcomings at protecting databases, trade secret protection is unlikely to be the solution to the problems faced by digital database makers.”).

93. See Human Genome Project, U.S. Dep’t of Energy Office of Sci., Summary of Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing, http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml#1 (last visited Jan. 10, 2007).

representatives of labs participating in publicly funded human genome sequencing when they held the first International Strategy Meeting on Human Genome Sequencing in Bermuda.⁹⁴ According to the Bermuda Statement, “all human genomic sequence information, generated by centres funded for large scale human sequencing, should be freely available and in the public domain in order to encourage research and development and to maximise its benefit to society.”⁹⁵ Organizations such as the National Institutes of Health’s National Human Genome Research Institute, the Department of Energy, and the Wellcome Trust require adherence to this policy as a condition of receiving funding.⁹⁶

According to its proponents, the Bermuda Statement data release policy not only fosters independent checking of the sequence data by other researchers,⁹⁷ but also prevents publicly funded large-scale sequencing centers from “establishing a privileged position in the exploitation and control of human sequence information.”⁹⁸ What is more, the policy also delivered a symbolic message that “the genome belongs to everybody.”⁹⁹ When sequencers met again in 2003 in Florida, they reaffirmed their commitment to the Bermuda Statement.¹⁰⁰

94. See *id.* (stating that the participants included “officers from, and scientists supported by, the Wellcome Trust, the UK Medical Research Council, the NIH NCHGR (National Center for Human Genome Research), the DOE (U.S. Department of Energy), the German Human Genome Programme, the European Commission, HUGO (Human Genome Organisation) and the Human Genome Project of Japan”); see also David R. Bentley, *Genomic Sequence Information Should Be Released Immediately and Freely in the Public Domain*, 274 SCI. 533, 534 n.1 (1996) (stating that participants in the first International Strategy Meeting on Human Genome Sequencing “included representatives of laboratories involved in human genome sequencing and of funding agencies, who met to discuss strategy, progress and plans, policies for data release, and the implications of such policies” and that “[t]he ‘Bermuda statement’ was endorsed unanimously by all participants”).

95. Human Genome Project, *supra* note 93; see also Bentley, *supra* note 94, at 534 n.1 (illustrating the large amount of institutional support for the Bermuda Statement).

96. See Lee Rowen et al., *Publication Rights in the Era of Open Data Release Policies*, 289 SCI. 1881, 1881 (2000); see also *infra* note 141 and accompanying text.

97. See Bentley, *supra* note 94, at 533 (emphasizing that scientific progress flourishes in an atmosphere of cooperation among researchers, which is best achieved by making the data publicly available).

98. Human Genome Project, *supra* note 93.

99. Eliot Marshall, *Bermuda Rules: Community Spirit, with Teeth*, 291 SCI. 1192, 1192 (2001) (internal quotation marks omitted) (quoting Ari Patrinos, Director of the U.S. Department of Energy’s office that funds genome research).

100. WELLCOME TRUST, SHARING DATA FROM LARGE-SCALE BIOLOGICAL RESEARCH PROJECTS: A SYSTEM OF TRIPARTITE RESPONSIBILITY 2 (2003), available at <http://www.wellcome.ac.uk/assets/wtd003207.pdf> (noting that “[t]he meeting attendees enthusiastically reaffirmed the 1996 Bermuda Principles”). In the report from this meeting, the participants designated the International HapMap Project, among other

While applauded by many in the research community,¹⁰¹ the Bermuda principles have not been met with universal favor. For example, a former Merck executive criticized the ease with which commercial researchers can engage in parasitic patenting.¹⁰² Certainly, those concerned with the advancement of science must be mindful that there is a danger in prioritizing prompt public access to genomic data without taking sufficient measures to prevent users of that data from filing patent claims that block others' use of the public data.¹⁰³

The drawback of putting data so freely in the public domain is illustrated by the Human Genome Project (HGP), a publicly funded effort to sequence the human genome.¹⁰⁴ In 1998, Craig Venter, at first a member of the International Human Genome Sequencing Consortium, later created a private company, Celera Genomics, which raced to complete a sequence of the human genome before the public project.¹⁰⁵ Ultimately, when the rival groups published their respective results around the same time, many observers alleged that Celera had achieved its goal by copying the Consortium's results, a charge Celera denied.¹⁰⁶ One commentator observed, "Without going into the detail of these controversial claims and counter-claims, it is clear that the race itself worried some public-sector researchers about the possible abuse of publicly available information that could be used later

research endeavors, as a "community resource project," meaning 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," and emphasized that "[t]he scientific community will best be served if the results of community resource projects are made immediately available for free and unrestricted use by the scientific community." *Id.* at 2-3.

101. See Marshall, *supra* note 99, at 1192 (citing Francis Collins, Director of the National Human Genome Research Institute (NHGRI) at the NIH, as claiming that the Bermuda Statement policy has "enabled the identification of more than 30 disease genes").

102. See *id.* (commenting that the research sponsors of the genetic databases should have taken steps to prevent others from using the database's resources to obtain private patents that inhibit the flow of information and hinder development).

103. See Andres Guadamuz González, Open Science: Open Source Software Licences and Scientific Research 20 (2005), available at <http://ssrn.com/abstract=764064> (observing that release of data into the public domain "is extremely useful for future researchers, but it does little to curb the further commercialization of the data").

104. See Nat'l Human Genome Research Inst., U.S. Dep't of Health & Human Servs., Human Genome Project Completion: Frequently Asked Questions, <http://www.genome.gov/11006943> (last visited Jan. 10, 2007) (describing how the genome data was made public, and the subsequent patent filings by private companies which may or may not limit the commercial use of the data).

105. See John Sulston, *Intellectual Property and the Human Genome*, in GLOBAL INTELLECTUAL PROPERTY RIGHTS: KNOWLEDGE, ACCESS AND DEVELOPMENT 61, 64-65 (Peter Drahos & Ruth Mayne eds., 2002).

106. See Guadamuz González, *supra* note 103, at 10.

on to make broad patent claims and commodify the biological data offered.”¹⁰⁷ “This fear seemed to be corroborated by the facts that by 1999, Celera had applied for the patenting of 6,500 human gene sequences, and by 2000 it had been awarded 300 patents.”¹⁰⁸

According to Professor Eisenberg, “the public sector Human Genome Project has paid a price for this policy, which has advanced the competitive position of their private sector rivals in the race to complete the sequence of the human genome and may have enhanced their patent positions as well.”¹⁰⁹ Likewise, though there was a public outcry when the publicly funded National Institutes of Health (NIH) first announced its decision to pursue patent rights in the first few thousand expressed sequence tags (ESTs) identified by Dr. Craig Venter when he was with NIH,¹¹⁰ ironically the private sector stepped in and obtained EST patents, effectively blocking others’ use of this sequence information.¹¹¹

Certainly, if the goals are, as expressed by one representative of the Sanger Center at the Wellcome Trust, to have “the value of the sequence [data] increase” by incorporating additional information from other sources, “encourage exploitation by a maximum number of commercial and academic centers that are keen to compete in the development of new therapeutic agents,” and maintain a nonexclusive environment,¹¹² the best way to achieve this is not a release into the public domain, but rather with an open source licensing scheme that requires users of the data to grant back to the Consortium all newly revealed haplotypes and haplotype tag SNPs so that all other database users would have nonexclusive freedom to use this information as tools for their own research. This strategy would minimize, and perhaps even eliminate, parasitic patenting, while enhancing the research tools developed by the

107. *Id.* at 10–11.

108. *Id.* at 11.

109. Rebecca S. Eisenberg, Correspondence, *The Promise and Perils of Strategic Publication to Create Prior Art: A Response to Professor Parchomovsky*, 98 MICH. L. REV. 2358, 2369 (2000); *see also* Eisenberg, *supra* note 10, at 73 (making similar comments).

110. For a brief description of expressed sequence tags (ESTs) and this controversy surrounding them in the early 1990s, *see* Donna M. Gitter, *International Conflicts Over Patenting Human DNA Sequences in the United States and the European Union: An Argument for Compulsory Licensing and a Fair-Use Exemption*, 76 N.Y.U. L. REV. 1623, 1654 n.207 (2001).

111. *See* Eisenberg, *supra* note 71, at 563 (noting that “the private sector took over the enterprise of cDNA sequencing”).

112. Bentley, *supra* note 94, at 534 (advocating the free release of genomic sequence data into the public domain).

HapMap database through contributions from numerous participants in the public, for-profit, and nonprofit sectors.

However, this approach certainly will not garner approval from the entire research community. Most prominently, 2002 Nobel Prize winner Sir John Sulston, who was also the first Director of the Sanger Center at the Wellcome Trust, which made the UK's contribution to the international HGP,¹¹³ strongly opposes the notion of licensing scientific works intended for public dissemination.¹¹⁴

2. *Little Harm to Scientific Progress Is Likely if the Enhanced Open Source Licensing Scheme Proposed Here Were to Place Genome Sequencing Centers in a Privileged Position.* As noted above, one important principle underlying the Bermuda Statement is “to prevent such [sequencing] centres establishing a privileged position in the exploitation and control of human sequence information.”¹¹⁵ Because the licensing system proposed here imposes both licensing fees and also transaction costs¹¹⁶ on researchers seeking to use the HapMap data, it is reasonable to assume that scientists working at the sequencing centers will enjoy an advantage by virtue of their free and open access to this data.¹¹⁷

This concern, while a valid one, is allayed somewhat by the fact that noncommercial researchers would enjoy reduced fees in order to facilitate their access to this data.¹¹⁸ Moreover, while a

113. Press Release, Wellcome Trust Sanger Inst., Sir John Sulston Awarded the 2002 Nobel Prize for Physiology or Medicine (Oct. 7, 2002), available at <http://www.sanger.ac.uk/Info/Press/2002/021007.shtml>.

114. See Sulston, *supra* note 105, at 64. In 2002, Sir Sulston coauthored *The Common Thread: A Story of Science, Politics, Ethics and the Human Genome*, in which he declared that, “The only reasonable way of dealing with the human genome sequence is to say that it belongs to us all—it is the common heritage of humankind.” JOHN SULSTON & GEORGINA FERRY, *THE COMMON THREAD: A STORY OF SCIENCE, POLITICS, ETHICS AND THE HUMAN GENOME* 113 (2002).

115. See Human Genome Project, U.S. Dep’t of Energy Office of Sci. Policies on Release of Human Genomic Sequence Data, http://www.ornl.gov/sci/techresources/human_genome/research/bermuda.shtml (last visited Jan. 10, 2007).

116. One would expect such transaction costs to remain relatively minimal, however, as generally publicly funded databases are relatively well-known in the scientific community, obviating the need for users to locate the owners, and licenses would be preestablished for each category of institution (e.g., pharmaceutical firm, nonprofit educational institution), rather than individually negotiated on a case-by-case basis.

117. It should be noted that this circumstance is not viewed negatively in all research fields. In the space sciences, there is actually a mechanism to reward producers of information that all share. See Rowen et al., *supra* note 96, at 1881. Such researchers typically enjoy special access to the instruments they develop and a proprietary period of time to analyze the data and publish their results. *Id.*

118. See *supra* note 68 and accompanying text.

nonexclusive licensing scheme clearly does impose financial and transaction costs upon researchers who were not involved in producing the data,¹¹⁹ this is preferable to a system where parasitic patenting is permitted to occur. A proliferation of private patent rights in such data would increase transaction costs even more so, both because parties desiring access to data covered by a patent would have to negotiate separately with each right holder and also because each patent holder that has the last patent required to practice a particular technology would have an incentive to hold out for exorbitant licensing fees.¹²⁰

Other nations that have developed national genetic databases, including Estonia and the United Kingdom, have employed a licensing approach.¹²¹ In 2001, the Government of the Republic of Estonia established the Estonian Genome Project Foundation (EGPF), called Eesti Geenivaramu in Estonian.¹²² This nonprofit organization aims “to create a database of health, genealogy and genome data that would comprise a large part of the Estonian population.”¹²³

In order to pursue commercial development, the EGPF partnered with EGeen Inc., a for-profit privately-held U.S. corporation based in the San Francisco area.¹²⁴ According to a representative of the EGPF, “[i]n exchange for funding and granting the [EGPF] a stake in the company, EGeen was granted exclusive commercial access to all data emerging from the Estonian Genome Project” from the inception of the project.¹²⁵ By mutual consent of the parties, the EGPF and EGeen terminated their exclusive access agreement in December 2004, such that EGeen was “no longer obliged to finance the activities of the Estonian Genome Project” and “[f]urther cooperation with EGeen will be taking place on equal bases with other researchers.”¹²⁶

119. See *supra* note 116 and accompanying text.

120. Cf. JEANNE CLARK ET AL., PATENT POOLS: A SOLUTION TO THE PROBLEM OF ACCESS IN BIOTECHNOLOGY PATENTS? 8–9, (2000), available at <http://www.uspto.gov/web/offices/pac/dapp/opla/patentpool.pdf> (describing some of the advantages of patent pools in the biotechnology context).

121. For a more detailed discussion of these databases, see generally Alice Hsieh, *A Nation's Genes for a Cure to Cancer: Evolving Ethical, Social and Legal Issues Regarding Population Genetic Databases*, 37 COLUM. J.L. & SOC. PROBS. 359, 388–96 (2004).

122. Eesti Geenivaramu, Estonian Genome Project: General Information, <http://www.geenivaramu.ee/index.php?lang=eng&sub=58> (last visited Jan. 10, 2007).

123. *Id.*

124. See EGeen, <http://www.egeeninc.com> (last visited Jan. 10, 2007).

125. E-mail from Mirjam Jalak, Assistant of the Board, Estonian Genome Project Found., to Donna M. Gitter, Assistant Professor of Legal and Ethical Studies, Fordham Univ. Sch. of Bus. (Dec. 13, 2005 3:14 EST).

126. *Id.*

Presently, the EGPF “owns and manages the database, negotiating non-exclusive license agreements with pharmaceutical and biotechnology companies while providing access to Estonian academics for free, or for a nominal charge.”¹²⁷ According to the license agreement, researchers must “provide the Estonian Genome Project with research results obtained by using the data received from the database. The research results shall supplement the genetic database of the Estonian Genome Project.”¹²⁸ The legislation creating the EGPF, the 2000 Estonian Human Genes Research Act (HGRA),¹²⁹ distinguishes between three different levels of data from research: while the first two, tissue and basic genetic data, are owned by the EGPF, “elaborate” genetic data created by the research are not regulated by the HGRA, and researchers and companies are therefore free to apply for patents on their findings.¹³⁰

Another national genetic database, the United Kingdom Biobank, which is still in its initial phase,¹³¹ will pursue a nonexclusive licensing approach from the outset.¹³² According to its website, “UK Biobank is a long-term project aimed at building a comprehensive resource for medical researchers” and will include blood and urine samples as well as answers to “a confidential lifestyle questionnaire” from about 500,000

127. See Hsieh, *supra* note 121, at 370; see also Estonian Genome Project Found., Data Available for Scientific Research, http://www.geenivaramu.ee/mp3/Issuance_of_the_data.pdf [hereinafter EGP Data] (describing the process of applying for and receiving permission to use the data). In order for non-Estonian researchers to use the data, they must be working in cooperation with Estonian scientists. See Estonian Genome Project Found., Estonian Genome Project in Short, http://www.geenivaramu.ee/mp3/EGP_%20in_short.pdf (last visited Jan. 10, 2007).

128. EGP Data, *supra* note 127.

129. See ESTONIAN GENOME PROJECT FOUND., ESTONIAN GENOME PROJECT 5, available at <http://www.geenivaramu.ee/mp3/trykisENG.pdf> (chronicling the passage of the Human Genes Research Act (HGRA) and the establishment of the Estonian Genome Project Foundation (EGPF) under the Act).

130. See Holger Breithaupt, *Pioneers in Medicine*, 4 EMBO REP. 1019, 1020 (2003). According to a lawyer involved in drafting the HGRA, this legislation “functions as a bridge between the collection of materials and patents.” *Id.* (quoting Tarmo Sild, an attorney in Tallinn, Estonia involved in drafting the HGRA).

131. According to Dr. Tim Peakman, Executive Director of UK Biobank, as of July 2006, the project was still in the peer review stage and was beginning to recruit participants. Telephone Interview with Dr. Tim Peakman, Executive Dir., UK Biobank (June 30, 2006). An initial pilot project included three thousand people, and over the next three and a half years the goal was to recruit a total of five hundred thousand participants. *Id.* Dr. Peakman hoped that U.K. Biobank would be available as a resource for researchers in about seven and a half years. *Id.*

132. See UK BIOBANK, POLICY ON INTELLECTUAL PROPERTY (“IP”) AND ACCESS 5 (2005), <http://www.ukbiobank.ac.uk/docs/UKBiobankIPandAccesspolicyfirstpublicdraft11.1.5final2.pdf> [hereinafter UK Biobank Policy] (explaining that access to the Biobank will be granted to meritorious applicants and will not be exclusive).

volunteers in the U.K.¹³³ UK Biobank will charge fees for access to its data, stipulating that “[f]ees for commercial use will generally be higher than those for non-commercial use, although consideration will be given to the impact of this on the full range of potential users and uses, including for example, on smaller companies or on innovative use in large companies.”¹³⁴

UK Biobank clearly seeks to encourage commercialization, specifying that it will permit users of the database to retain the rights to any intellectual property arising out of research using the Biobank¹³⁵ and “will not generally claim a share of any royalties generated from the results of research conducted using the [database].”¹³⁶ UK Biobank’s draft Policy on Intellectual Property and Access ensures, however, that licensees cannot use their future patents based on the licensed material “to restrict research use of such material by UK Biobank or its users.”¹³⁷ One proposed method for dealing with noncompliance is restriction on future access to UK Biobank.¹³⁸ In addition, as explained by UK Biobank’s Executive Director, Dr. Tim Peakman, if researchers fail to adhere to UK Biobank policy, they risk denial of access to funding by the U.K.’s Medical Research Council and the Wellcome Trust.¹³⁹

133. UK Biobank, Overview, <http://www.ukbiobank.ac.uk/about/overview.php> (last visited Jan. 10, 2007). Eventually, the Biobank will include “data derived from analyses of samples” and “data, materials and results from investigations conducted using” the Biobank. UK Biobank Policy, *supra* note 132, at 2.

134. UK Biobank Policy, *supra* note 132, at 7.

135. *Id.* at 3. UK Biobank reserves step-in rights to protect or exploit intellectual property, however, in cases where patent holders choose not to exercise their rights to do so. *Id.* at 4.

136. *Id.* at 8.

137. *Id.* at 7.

138. *Id.*

139. Telephone Interview with Dr. Tim Peakman, *supra* note 131. The Medical Research Council (MRC) is analogous to the U.S. National Institutes of Health. *See* Press Release, Science Media Centre, Scientists and Patients’ Groups Welcome New Breakthrough in Stem Cell Research (Aug. 13, 2003), *available at* http://www.sciencemediacentre.org/press_releases/03-08-13_stemcell.htm (“The Medical Research Council is a national organisation funded by the UK taxpayer. We promote research into all areas of medical and related science with the aims of improving the health and quality of life of the UK public and contributing to the wealth of the nation.”); *see also* Med. Research Council, About Us, <http://www.mrc.ac.uk/AboutUs/index.htm> (last visited Jan. 10, 2006). The Wellcome Trust, an independent charitable organization based in the U.K., is that nation’s “largest non-governmental source of funds for biomedical research.” Wellcome Trust, About Us, <http://www.wellcome.ac.uk/aboutus/> (last visited Jan. 10, 2007). It funds research not only in the U.K., but around the world, and describes itself as “the most diverse biomedical research charity in the world, supporting a spectrum of activity from basic science to history of medicine.” Wellcome Trust, Introduction, <http://www.wellcome.ac.uk/node3730.html> (last visited Jan. 10, 2007).

These enforcement strategies of UK Biobank could prove useful to creators of large-scale, publicly funded genomic databases. Denial of access for those who violate database access policies, as proposed herein,¹⁴⁰ is a useful approach vis-à-vis all types of researchers. Denial of NIH funding, however, holds promise mainly with respect to researchers who do not have access to fully private funding and rely, in whole or in part, on government funds. Such an approach already comports with current NIH policy, instituted in October 2003, which states that “investigators submitting an NIH application seeking \$500,000 or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible.”¹⁴¹ NIH currently subscribes to the Bermuda principles, however, and thus NIH data-sharing policy with respect to large-scale genome sequence data would necessitate the changes described in this Article in order to implement the approach described herein.¹⁴²

UK Biobank also contemplates a grantback policy, requiring users to provide the database with a copy of their research results so that other Biobank users can use the data without charge (other than the usual access fees all Biobank users pay).¹⁴³ This policy is mandatory “even if the data concerned are the subject matter of a patent.”¹⁴⁴ Thus, both Estonia and the U.K. have implemented aspects of the approach proposed in this Article. In each instance, a central government body offers nonexclusive access to a genetic database and charges user fees. In both cases, users are free to patent their discoveries but are also required to grant back to the database research results that will enrich the data available to other users. Other nations evidently see promise in such strategies, and their experience will offer useful case studies for the model developed in this Article.

140. See *supra* notes 77–78 and accompanying text.

141. Nat'l Insts. of Health, Office of Extramural Research, Final NIH Statement on Sharing Research Data (Feb. 26, 2003), <http://grants2.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.

142. See *supra* notes 93–96 and accompanying text.

143. UK Biobank Policy, *supra* note 132, at 8.

144. *Id.* Under U.K. law, by filing a patent application, the creator of the data can protect it vis-à-vis other Biobank users. *Cf.* Human Genome Project Information, U.S. Dep't of Energy Office of Sci., Genetics and Patenting, http://www.ornl.gov/sci/techresources/Human_Genome/elsi/patents.shtml (last visited Jan. 10, 2007) (describing all nations other than the United States and Philippines as “first inventor to file” jurisdictions).

3. *There Is a Market for Nonexclusive Licenses for Publicly Funded Genetic Databases.* Available evidence supports the argument that pharmaceutical firms would be willing to pay for access to databases such as the HapMap database. For example, by 1996, the publicly held corporation Incyte Pharmaceuticals, Inc. had succeeded in signing up as nonexclusive subscribers to its database eight pharmaceutical firms that “represented 28%, or \$8 billion, of the 1994 healthcare R&D spending by the top 50 pharmaceutical companies.”¹⁴⁵

According to former Incyte General Counsel Lee Bendekgey, however, “research dollars are much scarcer, certainly in large pharmaceutical companies” now than when Incyte established its database.¹⁴⁶ Thus, he sees reason to doubt that “many companies would be willing to pay more than a nominal amount for access to these types of databases,” noting that “[i]n general the pharmaceutical industry is happiest when basic research is conducted by the public sector (the government and academia) and placed in the public domain, so that they can access it for free.”¹⁴⁷

While pharmaceutical firms no doubt prefer to draw upon a rich public domain of genomic data, it is reasonable to think that, were the public sector and academia to cease placing genomic data in the public domain, a fee-based market would develop for such data. The financial value of such data is beyond dispute, as over \$100 million dollars (a large proportion of that U.S. tax dollars) was spent upon its collection and dissemination.¹⁴⁸ With reference to Celera Genomics, the company that raced the

145. *Incyte Signs Eighth Genomic Database Agreement with Zeneca; Launches Microbial Database Consortium*, PR NEWSWIRE, June 17, 1996 (on file with Houston Law Review). These firms included Abbott Laboratories, Hoechst AG, Hoffmann-La Roche, Johnson & Johnson, Novo Nordisk A/S, Pfizer Inc., Pharmacia & Upjohn, Inc., and Zeneca Ltd. *Id.* Incyte’s rival, Human Genome Sciences, began in 1993 by offering exclusive access to its EST database to SmithKline Beecham, but by 1996 had “saturated SmithKline with [drug-target] opportunities” and with SmithKline’s assent began to offer its EST database to three other pharmaceutical companies. Jon Cohen, *The Genomics Gamble*, 275 SCI. 767, 771 (1997) (alteration in original).

146. E-mail from Lee Bendekgey, Senior Vice President & Gen. Counsel, Nuvelo, Inc., to Donna M. Gitter, Assistant Professor of Legal & Ethical Studies, Fordham Univ. Sch. of Bus. (June 8, 2006 19:51 EST).

147. *Id.*

148. See Rai, *supra* note 6, at 147 (stating that, with respect to genomic databases, given “the high cost associated with generating initial data, and the corresponding value associated with such data, public funding of databases probably undermines the ability of private businesses to form around databases”); Nicholas Wade, *Gene-Mappers Take New Aim at Diseases*, N.Y. TIMES, Oct. 30, 2003, at A23 (stating that \$39 million of the total \$100 million cost of the HapMap Project came from the NIH).

publicly funded HGP to complete a map of the human genome,¹⁴⁹ Professor Rai has noted that “the availability of the public data placed a ceiling on what Celera could charge,”¹⁵⁰ which suggests that a fee-based market could develop in the absence of public domain databases funded by public dollars.

Though not of this view, Mr. Bendekgey nevertheless does think it feasible for publicly funded databases such as the HapMap Project to charge pharmaceutical firms a nominal fee to cover transaction costs.¹⁵¹ As observed by Mr. Bendekgey, such a fee, while not sufficient to fund the entire genomic project, would be one method of supporting a trade secret claim.¹⁵² This would enhance the protection of the data, as advocated in this Article.

There is a further reason that pharmaceutical firms might be willing to fund large-scale genomic projects. Such firms, which are downstream users of genomic data, are interested in having genomic information widely available, thereby generating the prior art that will “thwart the ability of upstream developers to obtain patents on the information.”¹⁵³ For example, beginning in 1999, a group of ten pharmaceutical firms and the Wellcome Trust, collectively known as the SNP Consortium, Ltd.,¹⁵⁴ began mapping SNPs¹⁵⁵ and ultimately succeeded in placing 1.8 million SNPs in the public domain, at a cost to members of about \$45 million.¹⁵⁶

Professor Merges has explained that the SNP Consortium was “intent on preempting the emerging anticommons problem.”¹⁵⁷ These firms could achieve the same goal by paying

149. See *supra* notes 105–08 and accompanying text.

150. Rai, *supra* note 6, at 161.

151. E-mail from Lee Bendekgey, Senior Vice President & Gen. Counsel, Nuvelo, Inc., to Donna M. Gitter, Assistant Professor of Legal & Ethical Studies, Fordham Univ. Sch. of Bus. (June 6, 2006 18:09 EST). See *infra* Part IV.C.5 for a discussion of the fee structure proposed here for access to large-scale, publicly funded genomic databases.

152. E-mail from Lee Bendekgey, *supra* note 146.

153. Rai, *supra* note 75, at 832–33 (citing Eisenberg, *supra* note 109, at 2358). See also *supra* note 75 and accompanying text.

154. For the members of the SNP Consortium, see Human Genome Project, U.S. Dep’t of Energy Office of Sci., Who Are Members of the SNP Consortium?, http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml#whoare (last visited Jan. 10, 2007).

155. “Single nucleotide polymorphisms (SNPs) are common DNA sequence variations among individuals and have great significance for biomedical research.” Atlas of Genetics and Cytogenetics in Oncology and Haematology, The SNP Consortium Ltd., <http://atlasgeneticsoncology.org/extdef.html> (last visited Jan. 10, 2007).

156. Robert P. Merges, *A New Dynamism in the Public Domain*, 71 U. CHI. L. REV. 183, 190 (2004).

157. *Id.* The SNP Consortium acknowledged this rather openly, explaining that its “overall IP objective is to maximize the number of SNPs the [sic] (1) enter the public

for access to a genomic database such as the HapMap database, in effect subsidizing a government effort to keep the data freely available to all downstream users. Indeed, the SNP Consortium was a major funder of the HapMap Project.¹⁵⁸ The licensing scheme described here would provide such firms enhanced protection from the problem of parasitic patenting that can arise when large-scale genomic projects place their data in the public domain. As an incentive for these firms, an arrangement could be made whereby private funders who contribute over a certain sum would not be required to pay a licensing fee for the data.¹⁵⁹

4. *The HapMap Database Qualifies for Trade Secret Protection and the Secrecy of the Data Likely Can Be Maintained.* The HapMap database qualifies for trade secret protection, as have many commercial databases in the past.¹⁶⁰ The Uniform Trade Secrets Act (UTSA) defines a trade secret as:

information, including a formula, pattern, compilation, program, device, method, technique, or process, that:

- (i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and

domain at the earliest possible date, and (2) to be free of third-party encumbrances such that the map can be used by all without financial or other IP obligations.” Eisenberg, *supra* note 10, at 72 (internal quotation marks omitted); see also E-mail from Lee Bendekgey, *supra* note 146 (“Many believe that the motivation behind private funding of the SNP database was to preclude Celera, Incyte and others from making a business out of proprietary SNP databases.”).

Professor Eisenberg has also noted that putting SNPs in the public domain facilitates their “challenge and validation by the scientific community,” thereby increasing the likelihood that any pharmaceutical products the firms develop using these SNPs will win regulatory approval. See Eisenberg, *supra* note 10, at 71.

158. See HapMap Participating Groups, *supra* note 11 (listing the SNP Consortium as a funding agency).

159. Professor Malinowski raises the concern, however, that researchers, especially academics,

are very sloppy regarding reporting follow-on invention under [Bayh-Dole] (they report the first generation invention and then cannot be bothered . . . and sometimes are very sloppy about even reporting the first generation invention). I could see sponsors demanding all kinds of assurances, only to have a violation by a sloppy researcher/entity entangled in the bigger effort.

E-mail from Michael Malinowski, *supra* note 75. One solution for this problem would be to bar further use by researchers who fail to comply with the HapMap policy.

160. See Eisenberg, *supra* note 71, at 563 (noting that “owners have been able to exploit the databases commercially by controlling access to them, in effect using contracts and trade secrecy to protect their intellectual property”); Rai, *supra* note 75, at 821 n.33 (stating that numerous owners of genomic databases license such information as trade secrets).

(ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.¹⁶¹

Professor Sandeen has explored in depth the UTSA's requirements in her work.¹⁶²

The first prong of the UTSA definition of a trade secret, that the information is "not . . . generally known,"¹⁶³ emphasizes that secrecy of the information is required both at the time that the purportedly secret information is disclosed by the trade secret owner to the person to whom he communicates it, and later, when infringement is alleged to have occurred.¹⁶⁴ Implicit in this idea is the notion that the information must be novel in some way, though not in the sense of the novelty requirement for patent law.¹⁶⁵ The HapMap database can indeed satisfy this requirement, as the Consortium is compiling the information for the first time and the data is not at all well known even within the genomic industry, much less outside.¹⁶⁶

As for the second prong of the UTSA definition of a trade secret, that the information is "not . . . readily ascertainable,"¹⁶⁷ this mandates that the information is not "available in trade journals, reference books, or published materials" and also not easily reverse engineered.¹⁶⁸ The HapMap database is not readily ascertainable, as it is not available elsewhere and, indeed, was compiled for the first time by the HapMap Consortium.

The UTSA also requires as the third prong of its definition of a trade secret that the database owner make "reasonable" efforts to "maintain its secrecy."¹⁶⁹ As noted by Professor Sandeen, "[b]ecause trade secrets have no value unless they can be used

161. UNIF. TRADE SECRETS ACT § 1(4) (amended 1985), 14 U.L.A. 537 (2005). As of December 2006, all fifty states had enacted some version of the Uniform Trade Secrets Act. See Uniform Law Commissioners, Uniform Trade Secrets Act, http://www.nccusl.org/Update/uniformact_factsheets/uniformacts-fs-utsa.asp (last visited Jan. 10, 2007) (listing states that have adopted the UTSA and the version adopted by each state).

162. See generally Sharon K. Sandeen, *A Contract by Any Other Name is Still a Contract: Examining the Effectiveness of Trade Secret Clauses to Protect Databases*, 45 IDEA 119, 126–44 (2005) (discussing the limits and effects of the Trade Secrets Act).

163. UNIF. TRADE SECRETS ACT § 1(4)(i) (amended 1985), 14 U.L.A. 537 (2005).

164. See Sandeen, *supra* note 162, at 132–33.

165. See NIMMER, *supra* note 92, ¶ 3.04 ("The level of novelty required does not correspond to patent law standards. . . . A trade secret must involve novel information, but the threshold of inventiveness is substantially lower."). For a discussion of novelty in the U.S. patent law context, see Gitter, *supra* note 110, at 1637–43.

166. For more information regarding the value of the HapMap data, see *supra* note 24 and accompanying text.

167. UNIF. TRADE SECRETS ACT § 1(4)(i) (amended 1985), 14 U.L.A. 537 (2005).

168. UNIF. TRADE SECRETS ACT § 1 cmt. 5 (amended 1985), 14 U.L.A. 537 (2005).

169. UNIF. TRADE SECRETS ACT § 1(4)(ii) (amended 1985), 14 U.L.A. 537 (2005).

and said use often requires disclosure to others, numerous cases recognize that absolute secrecy is not required.”¹⁷⁰ Rather, “[a]ll that is needed for information to maintain its trade secret status is ‘relative’ or ‘substantial’ secrecy.”¹⁷¹ Naturally, wide distribution of a database such as the HapMap database over the Internet presents special problems, in light of the fact that the database owner voluntarily discloses its trade secret to a wide range of users.¹⁷² In the case of the HapMap database, implementation of the proper encryption measures and the requirement that licensees execute confidentiality agreements would help to ensure trade secret status.¹⁷³

Finally, the last prong of the UTSA definition of a trade secret requires that the information “derives independent economic value, actual or potential, from not being generally known to . . . other persons who can obtain economic value from its disclosure or use.”¹⁷⁴ According to Professor Sandeen, this factor suggests that users’ willingness to pay for the database must lie in the fact that it is kept secret, as opposed to its completeness or veracity.¹⁷⁵ The HapMap database satisfies this criterion, as the data, which concerns the DNA extracted from volunteers from four different populations,¹⁷⁶ derives value from

170. Sandeen, *supra* note 162, at 139.

171. *Id.* at 139 & n.83.

172. *See id.* at 140 (discussing the need for security measures to protect trade secrets in online databases). It should be noted, of course, that trade secret law developed before widespread use of the Internet to disseminate databases and that the law will impede productive economic efforts should it deny trade secret status to online databases. *See* Michael Risch, *Why Do We Have Trade Secrets*, 11 MARQ. INTELL. PROP. L. REV. (forthcoming 2007) (manuscript at 27–30, on file with the Houston Law Review) (discussing the various economic justifications for trade secrets).

173. *See supra* note 67 and accompanying text (stating that the owner of a trade secret is permitted to use and disclose his secret pursuant to express contractual restrictions). As noted by Professor Sandeen, however, while a confidentiality agreement “is some evidence of reasonable efforts, it is not determinative of the issue.” Sandeen, *supra* note 162, at 140.

174. UNIF. TRADE SECRETS ACT § 1(4)(i) (amended 1985), 14 U.L.A. 537 (2005).

175. *See* Sandeen, *supra* note 162, at 141–42. Pursuant to the *Restatement (First) of Torts*, however, a proper inquiry in a trade secret case is the value of the information to the database owner and his competitors. *See* RESTATEMENT (FIRST) OF TORTS § 757 (1939). Professor Sandeen notes that some courts have interpreted the UTSA similarly, but contends this is in error. *See* Sandeen, *supra* note 162, at 141 (“While some courts are confused about the economic value prong of trade secret law, the language of the UTSA” clearly indicates that the “issue presented is not whether the information has value in the abstract, but whether the information (1) derives value from its secrecy and (2) is of value ‘to others.’”).

176. *See supra* note 19 and accompanying text (stating that blood samples will be taken from Nigeria, Japan, China, and the United States).

the fact that it is not generally known to others and will prove useful in determining genetic associations with disease.¹⁷⁷

Notwithstanding the enactment of the UTSA in 1990, courts still cite approvingly to the definition of a trade secret set forth in the *Restatement (First) of Torts*.¹⁷⁸ While noting that a precise definition of a trade secret is not possible, the *Restatement (First) of Torts* lists the following among the factors to be considered in assessing trade secret status:

- (1) the extent to which the information is known outside of his business;
- (2) the extent to which it is known by

177. See *supra* note 24 and accompanying text (explaining the uses of the HapMap, including its use for studying genetic associations with disease).

178. See, e.g., *Softel, Inc. v. Dragon Med. & Sci. Commc'ns, Inc.*, 118 F.3d 955, 968 (2d Cir. 1997) (“New York generally looks to section 757 . . . for its definition of trade secret.”); *Phillips v. Frey*, 20 F.3d 623, 628 (5th Cir. 1994) (quoting the *Restatement (First) of Torts* and case law in defining “trade secret”); see also 1 MILGRIM, *supra* note 67, § 1.01[1] & n.3 (noting that the *Restatement* definition of a trade secret has been “cited approvingly in virtually every U.S. jurisdiction”). The *Restatement* provides as follows:

Definition of trade secret. A trade secret may consist of any formula, pattern, device or compilation of information which is used in one's business, and which gives him an opportunity to obtain an advantage over competitors who do not know or use it. It may be a formula for a chemical compound, a process of manufacturing, treating or preserving materials, a pattern for a machine or other device, or a list of customers. It differs from other secret information in a business (see § 759) in that it is not simply information as to single or ephemeral events in the conduct of the business, as, for example, the amount or other terms of a secret bid for a contract or the salary of certain employees, or the security investments made or contemplated, or the date fixed for the announcement of a new policy or for bringing out a new model or the like. A trade secret is a process or device for continuous use in the operation of the business. Generally it relates to the production of goods, as, for example, a machine or formula for the production of an article. It may, however, relate to the sale of goods or to other operations in the business, such as a code for determining discounts, rebates or other concessions in a price list or catalogue, or a list of specialized customers, or a method of bookkeeping or other office management.

RESTATEMENT (FIRST) OF TORTS § 757 cmt. b (1939).

There are two main differences between the UTSA and *Restatement* conceptions of a trade secret. First, by defining “misappropriation,” separately from “trade secret,” the UTSA makes clear that a defendant's wrongdoing in terms of revealing confidential information is not sufficient to establish misappropriation of a trade secret; the existence of a trade secret must be established as a threshold matter. See UNIF. TRADE SECRETS ACT § 1(2), 14 U.L.A. 529, 537 (2005); UNIF. TRADE SECRETS ACT § 1(4), 14 U.L.A. 529, 538 (2005); see also Sandeen, *supra* note 162, at 130 & n.43 (2005) (“In this sense, the ‘business ethics’ rationale for trade secret law is not based upon the immorality of copying *per se*, but on the immorality of copying something that is truly secret and that the plaintiff took adequate steps to protect.”). Second, the UTSA, unlike the *Restatement*, “does not provide protection for confidential business information that is not a trade secret.” See *id.* at 132. The idea behind this change from the common law was to avoid conflicts between state trade secret law and federal patent law. See *id.* at 132 & n.50 (citing drafting history in support of the conclusion that the UTSA's “elimination of protection for mere ‘business information’ was not accidental,” but instead “prompted by a concern that providing protection for information that did not meet the definition of a trade secret would create a conflict between trade secret law and federal patent law”).

employees and others involved in his business; (3) the extent of measures taken by him to guard the secrecy of the information; (4) the value of the information to him and to his competitors; (5) the amount of effort or money expended by him in developing the information; (6) the ease or difficulty with which the information could be properly acquired or duplicated by others.¹⁷⁹

Applying the *Restatement* factors to the HapMap database reveals that the database could indeed qualify for trade secret protection if carefully shielded from disclosure. With regard to the first and second factors, a novel scientific discovery such as the HapMap is not at all well known even within the genomic industry, much less outside, which is why the HapMap Consortium has marshaled such resources in order to develop this data.¹⁸⁰ In terms of the third factor, as stated above, implementation of the proper encryption measures and the requirement that licensees execute confidentiality agreements would help to ensure trade secret status.¹⁸¹ With regard to the fourth and fifth factors, the value of the HapMap data is considerable, as demonstrated by the resources expended to collect it.¹⁸² Finally, the HapMap data could not easily be acquired or duplicated by others, as evidenced by the fact that genomic sequencing requires the resources of large-scale, publicly funded projects.¹⁸³

179. RESTATEMENT (FIRST) OF TORTS § 757, cmt. b (1939).

180. See *supra* note 148 and accompanying text (discussing the costs associated with the HapMap Project); see also HapMap Participating Groups, *supra* note 11 (listing research groups involved in the HapMap Project).

181. See *supra* notes 67 and 172 and accompanying text.

182. See *supra* note 148 and accompanying text. As noted by Professor Sandeen, under the UTSA, the amount of time, money, and effort expended by a database creator should not affect whether that database garners trade secret status. See Sandeen, *supra* note 162, at 136–37. However, in jurisdictions where the *Restatement* still constitutes binding or persuasive authority in trade secret cases, courts continue to assess this factor. *Id.* at 135–36; see also NIMMER, *supra* note 92, § 3:9 (contending that the amount of money that a firm has expended in creating a trade secret “can be a positive indication that protection should be granted to the information”).

183. An example of one such large-scale, publicly funded effort prior to the HapMap Project was the HGP, a thirteen-year project coordinated by the U.S. Department of Energy and the National Institutes of Health and aided by the contributions of the Wellcome Trust in the United Kingdom as well as researchers from Japan, France, Germany, and China. See Human Genome Project, U.S. Dep’t of Energy Office of Sci., http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml (last visited Jan. 10, 2007). A private firm, Celera Genomics, Inc., did indeed compete with the HGP to sequence the genome and achieved its goal at approximately the same time. See Donna M. Gitter, *supra* note 110, at 1630 (stating that, “on June 26, 2000, two independent teams of scientists, one from the HGP and the other from Celera Genomics Group (Celera), jointly announced that, working separately, each had completed a working draft of the genome’s entire sequence” (footnote omitted)). Nonetheless, many charge that Celera would have

Courts applying the six *Restatement* factors have held numerous times that customer lists are protected under trade secret law, where “it would be difficult to duplicate a customer list because it reflected individual customer preferences.”¹⁸⁴ This is particularly true where the list required substantial time and money to compile and contained information “which could only be achieved through personal solicitation.”¹⁸⁵ The HapMap database is analogous to such a list, as it is even more difficult and costly to create and relies on personal solicitation in order to compile the information.

One potential problem with designating the HapMap data a trade secret, however, is that trade secret status is lost once the information becomes “generally known.”¹⁸⁶ For example, if certain users of the data were to put the data in the public domain then trade secret status would be lost.¹⁸⁷ In light of the threat of parasitic patenting raised by placing such data in the public domain, serious consideration of alternatives such as trade secret status is essential. Indeed, trade secrecy would provide stronger protection than the HapMap Public License; while the latter approach hinges completely on contract law,

been unable to sequence the genome without access to the public domain information. *See supra* notes 104–09 and accompanying text.

184. *N. Atl. Instruments, Inc. v. Haber*, 188 F.3d 38, 46 (2d Cir. 1999).

185. *Giffords Oil Co., Inc. v. Wild*, 483 N.Y.S.2d 104, 106 (App. Div. 1984).

186. *See* UNIF. TRADE SECRETS ACT § 1(4)(i), 14 U.L.A. 529, 538 (2005) (noting that a trade secret “derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use”); *see also* RESTATEMENT (FIRST) OF TORTS § 757 cmt. b (1939) (“The subject matter of a trade secret must be secret. Matters of public knowledge or of general knowledge in an industry cannot be appropriated by one as his secret. Matters which are completely disclosed by the goods which one markets cannot be his secret. Substantially, a trade secret is known only in the particular business in which it is used. It is not requisite that only the proprietor of the business know it. He may, without losing his protection, communicate it to employees involved in its use. He may likewise communicate it to others pledged to secrecy. Others may also know of it independently, as, for example, when they have discovered the process or formula by independent invention and are keeping it secret. Nevertheless, a substantial element of secrecy must exist, so that, except by the use of improper means, there would be difficulty in acquiring the information.”); 1 MILGRIM, *supra* note 67, § 1.03 (“Indispensable to an effective allegation of a trade secret is proof that the matter is, more or less, secret.” (footnote omitted)).

187. *See* Alexander K. Haas, *The Wellcome Trust’s Disclosures of Gene Sequence Data into the Public Domain & the Potential for Proprietary Rights in the Human Genome*, 16 BERKELEY TECH. L.J. 145, 162 (2001) (“Trade secret protection, while perhaps applicable to the kind of information that genome companies have as long as it is secret, would be destroyed as soon as an independent publicly accessible database is published.” (footnote omitted)). Indeed, even if the information were not known to the general public, “[i]f the principal persons who can obtain economic benefit from [the] information are aware of it, there is no trade secret.” Sandeen, *supra* note 162, at 133 (quoting UNIF. TRADE SECRETS ACT § 1(4)(i) cmt. 5, 14 U.L.A. 529, 538 (2005) (alteration in original)).

the former provides an independent statutory basis for enforcing the private agreement between licensor and licensee.¹⁸⁸ What is more, owners of commercial genetic sequence databases have successfully used trade secret law (along with contract law) in the past to protect their investments.¹⁸⁹ In addition, while independent creation can destroy trade secret status,¹⁹⁰ this situation would present no problem to the HapMap Consortium, which actually wants the data to be available to as many users as possible.

Because publicly funded projects such as the HapMap database typically intend to foster publication by licensees, both in order to disseminate the data and to promote its verification by other researchers, it is worthwhile to consider whether the trade secrecy approach advocated here, which would limit to some degree publication by licensees, is compatible with those goals. In this instance, it appears that trade secrecy would not interfere with publication much more than the initial HapMap Project Public Access License already did. Because the initial Public Access License barred licensees from sharing the data with others who had not accepted the terms of the license, licensees already faced somewhat restrictive terms of publication. Specifically, licensees could “not include in publications the data on genotypes of individual HapMap samples . . . obtained from the Genotype Database,” though they could “publish conclusions based on such data,” as long as they cited the HapMap database so that others could access it on the same terms.¹⁹¹ More generally, there is a trend among some scientific publications to permit publication even though the genomic data supporting the research is maintained in a private database.¹⁹² Thus, publication could

188. Trade secret status for the HapMap data seems even more promising when one considers that secrecy need be maintained for a relatively short time, just until the derivation of haplotype data would be considered obvious and therefore no longer patentable. *See supra* note 37 and accompanying text.

189. *See supra* note 160 and accompanying text.

190. *See* 1 MILGRIM, *supra* note 67, § 1.01[1] (“One who discovers another’s trade secret properly, as, for example, by inspection or analysis of the commercial product embodying the secret, or by independent invention, or by gift or purchase from the owner, is free to disclose it or use it in his own business without liability to the owner.”).

191. Wash. Univ. in St. Louis Sch. of Med., Terms and Conditions for Data Access, SNP Research Facility Public Access License, <http://snp.wustl.edu/data/terms-and-conditions.html> (last visited Jan. 10, 2007) (explaining that this license applies to the HapMap Genotype Database).

192. *See, e.g.*, Pete Moore, *Publication with a Pinch of Privatisation*, SCIENTIST, Apr. 4, 2002, <http://www.biomedcentral.com/news/20020404/04> (describing a controversial decision in 2001 by the journal *Science* to publish two articles even though “the genomic data underpinning the publications” were kept in private databases).

still occur even under the approach advocated here, with certain restrictions comparable to those instituted by the initial HapMap data access policy.

5. *The Enhanced Open Source Access Model Proposed Here Will Attract the Sort of Small, Distributed Contributions That Advance the Goals of Scientific Research.* The fee-based access model proposed here has the potential to encourage small, distributed contributions from a wide array of researchers, thereby improving the quality of the data. In the classic computing context, the strength of the open source model is believed to lie in the fact that it relies on contributions from a large group of programmers. For example, Linux, the archetypal open source computer program, developed through the efforts of a large number of decentralized programmers contributing their code voluntarily through the Internet.¹⁹³ The benefits of this method are that the programs develop to better suit users' needs through this incremental process of innovation, and also that these improvements occur more quickly when compared with the rate of innovation typical in the conventional software industry.¹⁹⁴ Thus, the open source computing model comports well with the traditional Mertonian model of open science, which favors

193. See Hope, *supra* note 4, at 71–73. But see Josh Lerner & Jean Tirole, *Some Simple Economics of Open Source*, [L] J. INDUS. ECON. 197, 206 (2002) (claiming that, based on their research, “the open source process is quite elitist” and “[i]mportant contributors are few”). Professors Lerner and Tirole cited one study of twenty-five million lines of open source code as demonstrating that more than three-quarters of the nearly thirteen thousand contributors had made only one contribution and only four percent had made more than five contributions, which might seem to suggest a wide distribution of contributors. *Id.* at 204. However, they also found that “the top decile of contributors accounted for fully 72% of the code” and “the top two deciles for 81%”. *Id.* at 204–05 & fig. 2.

194. See Hope, *supra* note 4, at 73–75; see also Yochai Benkler, *Coase's Penguin, or, Linux and The Nature of the Firm*, 112 YALE L.J. 369, 376–77 (2002) (“Peer production has an advantage over firms and markets because it allows larger groups of individuals to scour larger groups of resources in search of materials, projects, collaborations, and combinations than is possible for firms or individuals who function in markets.”).

sharing of data among many researchers,¹⁹⁵ in part to ensure cross-checking and independent validation of research results.¹⁹⁶

Essential for such a model is a detailed, tiered licensing scheme that ensures access for noncommercial researchers at reduced fees. Professor Cooper Dreyfuss, in her work advocating compulsory licensing and a research exemption for the use of certain patented biotechnological research tools, explores the problem of defining what this Article refers to as noncommercial researchers (which she terms “basic researchers”).¹⁹⁷ Her ideas apply to the use of genomic databases such as the HapMap database by noncommercial researchers.

Professor Cooper Dreyfuss notes that researchers in the nonprofit, academic, and public sectors are often “engaged in projects with mixed goals,” and consequently a definition of noncommercial researcher “that relied on the commercial nature of the activities would not be effective.”¹⁹⁸ Many of those truly engaged in noncommercial research have links to industry, thereby potentially disqualifying them from being considered noncommercial researchers, while at the same time many researchers working for nonprofit educational institutions actively pursue patents.¹⁹⁹ Professor Cooper Dreyfuss therefore proposes a system whereby noncommercial researchers could gain access to patented research tools by executing a voluntary waiver that would, among other things, require them to “refrain from patenting the discoveries made in the course of using the invention for which the waiver was sought.”²⁰⁰ Applied to the HapMap Project, this model would permit academic, nonprofit, or

195. See ROBERT K. MERTON, *THE SOCIOLOGY OF SCIENCE: THEORETICAL AND EMPIRICAL INVESTIGATIONS* 273–75 (Norman W. Storer ed., 1973) (stating that, at the time of writing, an important norm of scientific research is free exchange of information); see also *HapMap-Tested Genotyping Assays and Microarrays Used for Association Studies*, AFFYMETRIX MICROARRAY BULL., Nov. 2005, at 12, 15, available at <http://www.microarraybulletin.com/downloads/HapMap-AMB.pdf> (quoting Dr. Richard A. Gibbs, Director, Baylor College of Medicine Genome Sequencing Center, and Wofford Cain Professor, Department of Molecular and Human Genetics, Baylor College of Medicine, as stating that “[t]o a large extent, the measure of success of a large genome project should be its ability to enable the activities of individual researchers”).

196. See Eisenberg, *supra* note 10, at 71 (stating that putting SNPs in the public domain facilitates their “challenge and validation by the scientific community,” thereby increasing the likelihood that any pharmaceutical products the firms develop using these SNPs will win regulatory approval).

197. See Rochelle Cooper Dreyfuss, *Varying the Course in Patenting Genetic Material: A Counter-Proposal to Richard Epstein’s Steady Course* 9–11 (N.Y. Univ. Law School, Pub. Law Research Paper No. 59, 2003), available at http://papers.ssrn.com/abstract_id=394000.

198. *Id.* at 9.

199. *Id.*

200. *Id.*

public-sector researchers engaged in noncommercial activity to gain access to the HapMap data at a very low cost²⁰¹ so long as they agreed not to commercialize the results of their research should they succeed in associating a particular haplotype with a specific phenotype.²⁰²

As explained by Professor Cooper Dreyfuss, the waiver provision is unlikely to suffer abuse by unscrupulous commercial researchers feigning to be noncommercial actors because the requirement to refrain from patenting ensures that only those who are not seeking to commercialize their work are likely to pursue this option.²⁰³ A related advantage of the waiver approach is that it will stimulate research in areas that pure market forces would not because the potential for profit from such research is low.²⁰⁴

Of course, scientific research often leads to unexpected discoveries, meaning that a putatively noncommercial researcher might ultimately make a potentially profitable discovery and subsequently wish to patent and commercialize it. In such a case, Professor Cooper Dreyfuss proposes a “buyout” option whereby the researcher is permitted to pay for the use of the patented technology which it had previously been entitled to use for a greatly reduced fee.²⁰⁵ This provision is similarly applicable to a publicly funded genomic database such as the HapMap database, in that the HapMap Consortium could retroactively charge the formerly noncommercial user (now a commercial user) a higher fee for database access.

In terms of the fee charged to commercial database users, one possibility organizations might consider is “value-based

201. According to Dr. Tim Peakman, the issues surrounding pricing are complex and require resolution soon for the benefit of the providers, the researchers, and the participants. Telephone Interview with Dr. Tim Peakman, *supra* note 131. Because UK Biobank is a nonprofit organization, the fees charged to researchers will reflect this. *Id.*

202. All users of the HapMap database would also benefit to the extent that noncommercial researchers furnish grantbacks of their data to the HapMap database. *See supra* Part IV.A. This lessens the potential resistance of for-profit firms toward subsidizing the research of noncommercial entities through payment of HapMap licensing fees.

203. *See* Cooper Dreyfuss, *supra* note 197, at 10 (“[I]t is highly likely that the waiver provision will make the exemption attractive mainly to those whose work is truly basic—that is, not expected to yield readily commercialized output.”).

204. *Id.* (“The waiver approach, in other words, is partly attractive because of its capacity to identify the situations where a chill on research is a potential problem” because “inefficient ‘hunting’ is especially likely to occur.”).

205. *Id.* at 11 (“Because research is often serendipitous, it may sometimes be the case that a scientist who signed a waiver later makes a commercially significant discovery Because such an invention may require patent protection to promote and facilitate further development, buyout should be permitted.”).

pricing.”²⁰⁶ According to this approach, the price charge for database access would depend upon the stage of the research process, ranging, for example, from the phase of initial drug “discovery to post-marketing surveillance.”²⁰⁷ Thus, a researcher involved in the early stages of the drug discovery process would face a lower licensing fee in light of the high cost and low probability of success of his endeavor.²⁰⁸ In contrast, a researcher using the data to support a clinical trial encounters significantly less risk and a higher likelihood of reward and therefore would pay a correspondingly higher licensing fee.²⁰⁹ A researcher at the stage of clinical trials would be willing to pay the licensing fee in view of the fact that any acceleration of the research process would translate into an earlier market entry and thus higher profits.²¹⁰ The method proposed here of varying the fees charged by noncommercial and commercial users of the HapMap data would facilitate use by noncommercial researchers, while still fostering patentable innovations and garnering licensing fees from those who make such discoveries.

V. CONCLUSION

In order to preserve the accessibility of genomic data generated by the HapMap Project, the HapMap Consortium drew inspiration from the open source software movement, implementing a data access policy that conditioned access to the data upon users’ agreement not to restrict others’ use of that data. As explored in this Article, this version of an open source model suffers from several drawbacks in the biotechnology context and therefore runs the risk of permitting parasitic patenting.

First, the HapMap Project’s open source clickwrap data access policy gave the HapMap Consortium an action for breach of contract but did not actually preclude patenting by licensees who violated the policy. Second, the HapMap Project’s open

206. Telephone Interview with Dr. Tim Peakman, *supra* note 131. Dr. Peakman emphasized strongly the noncommercial nature of U.K. Biobank, while at the same time noting that a “sensible” pricing approach for commercial and noncommercial researchers is required to promote the goals of the project. *Id.*

207. *Id.*

208. According to one recent estimate, “it typically takes about 15 years and costs up to \$800 million to convert a promising new compound into a drug on the market.” Peter Gwynne & Gary Heebner, *Drug Discovery and Biotechnology Trends: Recent Developments in Drug Discovery: Improvements in Efficiency*, SCL, Feb. 7, 2003 (Special Advertising Supplement), available at http://www.sciencemag.org/products/ddbt_0207_Final.dtl.

209. Telephone Interview with Dr. Tim Peakman, *supra* note 131.

210. *Id.*

source data access policy did not bind third parties who obtained the data through means other than the HapMap website and therefore ran the risk that such third parties could freely violate the terms of the access policy. Third, the HapMap Project's open source data access policy lacked a clear enforcement mechanism and suitable remedy. The HapMap Consortium would have had to bring suit to enforce the data access agreement, which would strain the financial and administrative resources of this nonprofit project, and the Consortium also would not have been able to calculate with specificity the damages for breach of its contract in light of the nonprofit nature of the project. Finally, the HapMap Consortium likely would have been unable to enforce its clickwrap data access license internationally since such agreements are not enforceable worldwide.

This Article therefore proposes an enhanced open source model, suggesting implementation by large-scale genomic projects of a nonexclusive, nonroyalty-bearing licensing policy for their data. Projects such as the HapMap Consortium essentially could protect data as a trade secret, offering subscription agreements at varying rates for nonexclusive access to the database and requiring users to execute confidentiality agreements prohibiting them from revealing the data. In addition, users would owe a grantback, which in the case of the now-completed HapMap Project, would have meant that any entity that derived haplotypes and haplotype tag SNPs would grant back to the Consortium and to all other users of the database nonexclusive freedom to use these haplotypes and haplotype tag SNPs as tools for their own research.

The proposed enhanced open source licensing scheme, which relies on the doctrine of trade secrecy, has the potential to solve the problems arising from the original HapMap license agreement and could therefore prove useful for other large-scale genome sequencing endeavors. First, the genomic data, which typically gets released on an ongoing basis, would be encrypted and unavailable to those who fail to agree to the stated terms of its use. Second, under applicable trade secret law, the enhanced open source model proposed here would bind third parties who obtained the data through means other than the HapMap website. Third, the database creator could enforce its rights under this model and win a significant measure of damages if successful in litigation because the typical remedy in a trade secret action is based on defendant's ill-gotten profits. Finally, there is a strong likelihood that the license proposed herein could be enforced internationally because U.S. trade secret law is comparable to the trade secret law of many other nations.

2007]

OPEN SOURCE PARADOX

1521

The enhanced open source licensing system proposed here is feasible and possesses a strong likelihood of success if implemented by other large-scale genome sequencing projects. First, this scheme is reconcilable with international agreements favoring data release into the public domain. Second, little harm to scientific progress is likely if the enhanced open source licensing scheme proposed here were to place genome sequencing centers in a privileged position. Third, there is a market for nonexclusive licenses for publicly funded genetic databases. Fourth, the secrecy of the HapMap data likely can be maintained and, even if it is not, wide dissemination furthers the HapMap project's goals. Finally, the enhanced open source access model proposed here will attract the sort of small, distributed contributions that advance the goals of scientific research.

As ever more public dollars are spent upon large-scale, publicly funded genome sequencing, it is critical to fashion alternatives that can maintain the accessibility of the data to all users and provide effective protection against parasitic patenting. The model proposed here, which is analogous to approaches launched in the Estonia and the United Kingdom, offers one promising solution worthy of exploration.