AN INTEROPERABILITY PRINCIPLE FOR KNOWLEDGE CREATION AND GOVERNANCE: THE ROLE OF EMERGING INSTITUTIONS

John Wilbanks[2]

Carolina Rossini[3]

This is a working paper, part of the research conducted under the Industrial Cooperation Project at the Berkman Center at Harvard, and originally written for the MINDS conference on Strategic Responses to Globalization held from 3 November 2009 to 6 November 2009 in Rio de Janeiro, Brazil. This second version will be presented at the II International Conference of Portuguese Speaking Countries, which will be hold from 25 to 27 de February de 2010, in Lisbon, Portugal.

ABSTRACT

This study examines the relationships among funders, research institutions, and the "units" of knowledge creation and local knowledge governance, which are hosted inside research institutions. Our goal is to uncover the knowledge spaces where commons-based approaches, peer production, and modes of network-mediated innovation have - and have not - emerged, and to examine the conditions under which these approaches either flourish or are discouraged. Our rationale is that the emergence of novel, democratized, and distributed knowledge governance represent a meaningful complement to more traditional systems, with the potential to create new public knowledge goods accessible to a global civil society and spur innovation in previously unforeseen ways.

---

[2] John Wilbanks is Vice President for Science at Creative Commons.

[3] Carolina Rossini is a Research Fellow at the Berkman Center for Internet & Society at Harvard Law School, coordinating the Industrial Cooperation Project.

The first section of this paper is an introduction to distributed knowledge creation and open systems for knowledge transactions (including but not limited to copyrighted and patented knowledge-embedded[4] products). The second section contains a case study of the complex and interlocking system of relationships governing knowledge-embedded products in the field of genomics, as well as some experimental interventions to adjust these relationships with the goal of maximizing either the total knowledge output or the value captured from the knowledge products. Although we have focused our case study on genomics, which offers a rich set of varied knowledge products and cases for study, the rationale we present in this discussion paper is also applicable to a variety of areas, from educational resources to alternative energy related technology.

The third section examines in particular the role of national innovation policy and the specific relationships that research institutions have with public policy and governmental funders, including the role of government and policy mandates (either towards enclosure or openness) as regards knowledge-embedded products. We have chosen to focus on the role of universities because their upstream role in the innovation ecosystem can influence downstream governance of knowledge through creation of cooperation arrangements, the strategic retention of rights, and/or publication and distribution of knowledge products.

[4] "Embedded knowledge: is knowledge which resides in systemic routines. The notion of 'embeddedness' was introduced by Granovetter (1985), who proposed a theory of economic action that, he intended, would neither be heavily dependent on the notion of culture (i.e. be 'oversocialized') nor heavily dependent on theories of the market (i.e. be 'under-socialized'): his idea was that economic behaviour is intimately related to social and institutional arrangements. Following Badaracco (1991), the notion of embedded knowledge explores the significance of relationships and material resources. Embedded knowledge is analyzable in systems terms, in the relationships between, for example, technologies, roles, formal procedures, and emergent routines. This is how, for example, Nelson and Winter (1982) analyzed an organization's capabilities. They noted that an individual's skills are composed of sub-elements which become co-ordinated in a smooth execution of the overall performance, impressive in its speed and accuracy with conscious deliberation being confined to matters of overall importance; thus, they maintained, may an organization's skills be analyzed. In addition to the physical and mental factors that comprise individual skills however, organizational skills are made up of a complex mix of interpersonal, technological and socio-structural factors. Similar approaches include Levitt and March's (1988) development of the notion of organizational routines (which, they suggest, make the lessons of history accessible to subsequent organizational members) while other writers refer to 'organizational competencies' (Prahaled and Hamel 1990). A related orientation has been proposed by Henderson and Clark (1990) who distinguish between the knowledge of specialist elements in an organization ('component knowledge') and knowledge about how such elements interact ('architectural knowledge'); architectural knowledge is often submerged within an organization's taken-for-granted routines and interactions, yet is central to an understanding of its strengths and weaknesses." *Knowledge, knowledge work and organizations: an overview and interpretation* . Organization Studies, Mid-Winter, 1995 by Frank Blackler)

Finally, we make a series of provisional recommendations for policymakers, funders, and research institution leadership that might decrease transaction costs for knowledge products, increase the incentive for collaborative knowledge production and reuse, and increase the odds of novel methods of knowledge creation, distribution and transfer emerging from an increasingly connected world.

Introduction

Before the advent of the Internet and the Web, the study of knowledge and its creation was primarily the province of the epistemologist. From a pragmatic perspective, knowledge was created through the interactions of the professors and the students, the texts and the experiments, and was captured through the process of writing. A small number of these writings were preserved for posterity as journal articles, theses, or textbooks, and a smaller number again of these became classics of knowledge.[5] These classics were heavily cited, reprinted, translated, summarized in review articles, embedded into standard textbooks, and more.

However, the revolution of digital technologies and networked communications brings the concerns of the philosopher to a much broader audience. The modern governance of knowledge in institutions covers a complex ecosystem of government, funders, universities, professors, students, textbooks, network-accessed content, software, research tools, research methods, and more[6]. The idea of the "article" or the "book" or the "thesis" as a canonical unit of knowledge, preserved and re-printed without constant and often significant changes, is rapidly finding parallel competition by the ability of the network to facilitate rapid, distributed, and

---

[5] The Agony And The Ecstasy: The History And Meaning Of The Journal Impact Factor. Eugene Garfield, Fifth International Congress on Peer Review in Biomedical Publication, September 2005, retrieved at http://garfield.library.upenn.edu/papers/jifchicago2005.pdf

[6] Knowledge governance is a broad concept that embraces different forms of governance mechanisms influencing the production, appropriability and dissemination knowledge. As a provisional definition, the "knowledge governance approach" is characterized as a distinctive, emerging approach that cuts across the fields of knowledge management, organization studies, and innovation and competition policies. Knowledge governance is taken up with how the deployment of governance mechanisms influences knowledge processes, such as sharing, retaining and creating knowledge". "As an analytical perspective, it encompasses intellectual property rules and regulations but supersedes it by drawing on those fields and disciplines in order to identify the contours of the new knowledge ecology, and to support alternative governance mechanisms for organizational and business models which are emerging as complements - or alternatives - to the instituted intellectual property regime we now have (Burlamaqui 2009). From Intellectual Property to Knowledge Governance: An Evolutionary Perspective and its Implications for Development, *Leonardo Burlamaqui and Mario Cimoli, Jun 29, 2009*

incrementally created knowledge structures like wikis and free / *libre* open source software.[7]

Almost parallel to this emergence of ecosystem-based knowledge creation sits the research institution and the research funder. Institutions retain specific roles, rights and privileges that govern knowledge production, transfer, and exploitation. They are both generator of knowledge at scale, and purchaser of knowledge at scale (for instance, through library and textbooks acquisition). They are often subject to policy mandates of government, legislative commandments, and funder terms and conditions. Many institutions have evolved bureaucratic infrastructure for inbound and outbound knowledge governance via the technology transfer office and sponsored research office. But the institution and its infrastructure have existed to date in almost a separate world from the network-based, peer-produced-and-governed world.[8]

This infrastructure now exists inside a different information and innovation ecology driven by the Internet. The interconnection of institutions - companies, universities, NGOs, and more – creates a potential world in which a more standardized set of knowledge transactions yields great overall knowledge creation as well as financial opportunity. And in this new ecology, network-driven distribution of knowledge creation is only one challenge to institutions accustomed to playing key roles in the knowledge governance process. Other challenges include changes to technology transfer, intellectual property licensing, data policies, and faculty incentives. But universities in particular hold the promise of serving as "public spaces" in which individual knowledge creation via distributed systems is noted and rewarded, as well as engaging in purposeful and systematic knowledge distribution via both formal and informal transaction channels[9].

Governments and private funders of knowledge creation have begun to recognize these changes and to look for leverage points at which they might intervene in favor of more open and

---

[7] See: It's The Customer, Not The Container – debate at the Society for Scholarly Publishing's Scholarly Kitchen site. http://scholarlykitchen.sspnet.org/2009/09/24/john-wilbanks-its-the-customer-not-the-container/

[8] There are some very notable exceptions to this general rule, such as MIT's Open Courseware Initiative – see http://ocw.mit.edu. However, of all educational institutions, those exceptions form a very small percentage.

[9] It should be noted that there is significant movement by universities to adopt local policies and technologies, typically based on the library system, to facilitate public access to journal articles and e-theses developed on campus. See the Registry of Open Access Repositories at http://roar.eprints.org/ for a lengthy list of such university policies.

cooperative knowledge governance. Thus, we see movements in favor of open access of research outputs[10] and educational resources towards taxpayer funded research, private research foundations enforcing a sharing regime on research institutions in exchange for research funding[11]. However, we also see the impact of US pre-Internet legislation on knowledge governance systems across the world as countries eager to create national knowledge-driven innovation systems akin to that in the US, which can come into both real and perceived conflict with the sharing regimes under experimentation[12].

In this brief paper we attempt to “see the ocean in a grain of sand” by examining two new forms of knowledge governance (distributed-individual and open-institutional), performing a short case study of knowledge governance creation in the genomic-sciences sector, and looking at the role of the university in particular as a knowledge broker in the networked society.

1. Distributed Innovation, Open Innovation and Knowledge Governance

In a distributed, networked context, knowledge can be created without a central authority assigning tasks and without the maximalist approach to intellectual property associated with traditional forms of innovation diffusion and exploitation. Rather, the communities are formed by many different individuals, participating for very different sets of reasons and incentives[13], who self-organize around challenges and tasks. This type of knowledge creation has been named “distributed innovation”[14] and is most easily seen in peer-produced systems like the Wikipedia online encyclopedia and the Linux operating system, both of which are built for “free” by thousands of individuals, connected by the Internet.

Distributed innovation (DI) examines the power we see in Wikipedia and Linux development, in which a collected set of individuals each donate individual actions to a collective work which

[10] U.S. National Institutes of Health Public Access Policy, http://publicaccess.nih.gov/
[11] Wellcome Trust Open Access Policy, http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/index.htm
[12] Bayh-Dole alike legislation from Brazil in 2003 and 2004 and South Africa in 2009 and 2010 are clear examples.
[13] Moglen, E. Anarchism Triumphant: Free Software and the Death of Copyright. First Monday, 1999. Available at: http://emoglen.law.columbia.edu/my_pubs/anarchism.html
[14] e.g., Lakhani, Karim R., and Jill A. Panetta. "The Principles of Distributed Innovation." *Innovations: Technology, Governance, Globalization* 2, no. 3 (summer 2007).

then “snap together” into a coherent group through standard technical systems, digital networks[15], and a community-born set of norms and rules[16]. This is a novel ecosystem for knowledge creation and presents both challenges and opportunities for the traditional, print-based knowledge governance systems still common in the academy. DI brings a significant increase in individual empowerment to participate directly in knowledge governance through membership in multiple distributed knowledge communities as part of the move to network forms of knowledge product creation. These communities exist outside the traditional confines of the academy and are difficult to assign value to, and have yet to be addressed by most institutions which remain resolutely focused on traditional metrics of citation in “high impact” scholarly journals, patents acquired and licensed, and so forth.

However, as easy as it is to recognize distributed knowledge creation when we look at Wikipedia or GNU / Linux or some initiatives on open educational resources[17], it is incredibly difficult to “program” distributed innovation. Participants in the knowledge creation must agree on a standard method to publish knowledge, and there must be significant infrastructure to support the evolution, distribution, storage, search, retrieval, and verification of knowledge created. As this kind of agreement is hard to achieve, it is far easier to note examples of distributed knowledge creation than it is to spark the creation of a new one ex ante.

These requirements are not new to the network: all of them conform closely to the way that scholars have operated since the creation of the scholarly journal in 1665[18]. The article and the

[15] Wikipedia Meta Wiki, “Wiki Is Not Paper,” http://meta.wikimedia.org/wiki/Wiki_is_not_paper
[16] Similarly, you find that “[c]ooperation in peer-production processes is usually maintained by some combination of technical architecture, social norms, legal rules, and a technically backed hierarchy that is validated by social norms.” Benkler, Yochai, The Wealth of Networks. 2006. p.104
http://cyber.law.harvard.edu/wealth_of_networks/Download_PDFs_of_the_book
[17] Open educational resources are “teaching, learning and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use or re-purposing by others. In this sense, OER includes learning content, software tools to develop, use and distribute any kind of content, and implementation resources such as open licenses. From the Cape Town Open Education Declaration: ‘These resources include openly licensed course materials, lesson plans, textbooks, games, software and other materials that support teaching and learning. They contribute to making education more accessible, especially where money for learning materials is scarce. They also nourish the kind of participatory culture of learning, creating, sharing and cooperation that rapidly changing knowledge societies need.’” Rossini, Carolina. The State and Challenges of Open Educational Resources in Brazil. 2010. (p.17)
http://www.soros.org/initiatives/information/focus/access/articles_publications/publications/oer-brazil-20100101
[18] In 1662, the newly formed 'Royal Society of London for Improving Natural Knowledge' was granted a charter to publish by King Charles II and on 6 March 1665, the first issue of *Philosophical Transactions* was published under the visionary editorship of Henry Oldenburg, who was also the Secretary of the Society. The first volumes of what is

book are standard knowledge "containers" and are governed through a set of gating functions. Peer review verifies, publishers distribute, libraries store, retrieve, and serve out the knowledge in journals and books. Edits to the scholarly canon are made via new papers, corrigenda, and outright repudiation of previous studies[19]. In this context we can actually view scholarly publishing as a predecessor to the wiki approach – the culture of knowledge advancing, edit by edit.

On top of this innate synchrony with the concepts of distributed knowledge creation, institutions have tried to make elements of distributed innovation and digital systems part of their daily lives. The technical infrastructures inside the libraries have made the leap to digital portals, web-based catalogs, and more, but the new systems are primarily focused on the movement of the traditional units of knowledge in their digital forms rather than encouraging a rapid, decentralized knowledge governance process.[20] Component elements of peer produced systems often appear at universities spontaneously, or are funded by government research programs, but as yet have not linked up to transform the system at a broad level[21] outside of specific domains of research.

So why then has distributed innovation in knowledge creation not taken root in academics the way it has in encyclopedias and software? One key reason is the lingering effect of the technology of paper in which academic scholarship developed. [22]The print culture is embedded deeply into the academy and represents a major blocking agent to the adoption of distributed innovation. Rewards, incentives, and metrics for academic professionals are deeply tied to print-

---

now the world's oldest scientific journal in continuous publication were very different from today's journal, but in essence it served the same function; namely to inform the Fellows of the Society and other interested readers of the latest scientific discoveries. As such, *Philosophical Transactions* established the important principles of scientific priority and peer review, which have become the central foundations of scientific journals ever since.
From History of Philosophical Transactions of the Royal Society, http://rstl.royalsocietypublishing.org/

[19] A timely example of this would be the January 2010 decision by the Lancet, a venerable British journal, to withdraw a long-criticized and controversial study linking autism to vaccines.

[20] The Need for a Knowledge Web for Scholarship, Carolina Rossini, http://publius.cc/need_knowledge_web_scholarship/020509

[21] This is easiest to see in the life sciences, where the Open WetWare wiki system for laboratories and protocols has swept through the field of "synthetic biology" very quickly, and where government-funded efforts to create ontologies and open data systems is the international norm. Newer research disciplines like synthetic biology and disciplines with large, rare data generation centers (like high energy physics and astronomy) anecdotally appear to be more amenable to novel forms of knowledge governance.

[22] This insight comes from conversations by the authors with Jean-Claude Guédon

based metrics like citations, references, and impact factors. The existing systems of knowledge governance and credit allocation are not well aligned with a distributed knowledge creation environment, and the kind of authority rewarded in academia (typically resulting from award of advanced degrees) is not always the same kind of authority rewarded in a distributed knowledge system.

An institution or group of institutions wanting to enable true distributed knowledge governance would face a socio-technical set of challenges to implementation, from the difficulty of rewarding participation in peer production of knowledge, the difficulty of defining knowledge into forms that work on wikis and other new modes of knowledge creation and distribution, on the complexity of curating data and databases, and on the limitations of library capability in the long-term storage and preservation of data – to name but a few.

There is a key lesson for educational institutions aiming for distributed knowledge creation to draw from the technical world: the idea of interoperability. Interoperability allows unrelated systems to communicate and function in ways unexpected by their designers. The idea of interoperability as something that scales from technology to knowledge itself has emerged alongside the rise of the digital commons in culture and software. In this view, it is not only computer networks that must interoperate, but intellectual property rights and semantic understanding, so that distributed peer production of knowledge can make the leap from an encyclopedia into the sciences and other research disciplines.

Thus to expand the idea of interoperability from technology to knowledge, it is worth examining where the choice to "separate concerns" [23] in the early design of the internet itself enabled the

[23] "Let me try to explain to you, what to my taste is characteristic for all intelligent thinking. It is, that one is willing to study in depth an aspect of one's subject matter in isolation for the sake of its own consistency, all the time knowing that one is occupying oneself only with one of the aspects. We know that a program must be correct and we can study it from that viewpoint only; we also know that it should be efficient and we can study its efficiency on another day, so to speak. In another mood we may ask ourselves whether, and if so: why, the program is desirable. But nothing is gained – on the contrary! – by tackling these various aspects simultaneously. It is what I sometimes have called 'the separation of concerns', which, even if not perfectly possible, is yet the only available technique for effective ordering of one's thoughts, that I know of. This is what I mean by 'focusing one's attention upon some aspect': it does not mean ignoring the other aspects, it is just doing justice to the fact that from this aspect's point of view, the other is irrelevant. It is being one- and multiple-track minded simultaneously." Dijkstra, Edsger W. (1982), "On the role of scientific thought", in Dijkstra, Edsger W., *Selected writings on Computing: A Personal Perspective*, New York, NY, USA: Springer-Verlag New York, Inc., pp. 60–66, ISBN 0-387-90652-5

emergence of distributed innovation and knowledge construction. Separating concerns is a subtle concept, but a very powerful one. It runs through the worlds of distributed innovation and peer production, from the human level – edits to one article within Wikipedia do not affect other articles inside Wikipedia – to the technical level, in which one can simply install a piece of software on a computer without a deep knowledge of the levels of software already in operation.[24]

The technical separation of concerns[25] that allowed the creation of Wikipedia was the necessary first step towards the development of an online encyclopedia, but was not sufficient. Since Wikipedia is a collective creative work, it has requirements that fall outside the scope of the technical standard systems that form the network – in particular, copyright. Added to the stack of standards for technical knowledge creation are the foundational legal tools of copyleft[26] and open source software licenses[27], which guarantee the legal interoperability of the code on top of the technical interoperability. The invention of the public copyright license was essential to remove the legal uncertainty around the collective creation of software code: everyone's contributions not only snapped together technically, but worked through the shared license to snap together legally. Wikipedia rights were first secured through the GNU Free Documentation License and is

[24] The implementation of separation of concerns increases the "modularity" of software systems, technical networks, online encyclopedias, and more. The knowledge product is built into a cohesive whole out of a set of disparate parts or modules. This approach allows for many different knowledge creators to work on a similar task, without deep personal interaction, while still knowing that the knowledge created (whether wiki articles into an encyclopedia, or a web browser) will "snap together" if everyone involved follows the community and technical standards.

[25] In a cultural knowledge product like Wikipedia, the technical stack of infrastructure is embedded, and invisible to the end user. Wikipedia exists at one level because of the existence of the Internet itself, because the creator of wiki software did not have to ask permission of the creators of the Web for permission to run software on it, just as Tim Berners-Lee did not have to ask permission of the creators of the Internet to release the Web to run on it, and so on. The technical design of the Internet separates the issues of how to move information around from the issues of how to run applications and to present information. The system is designed from the beginning to allow for technologies that were not yet imaginable by the creators to flourish, for unintended uses to be enabled by default. Wikipedia sits atop this structure, technically, as do the end-user interfaces on users' computers, cell phones, and other network access points.

[26] "Copyleft uses copyright law, but flips it over to serve the opposite of its usual purpose: instead of a means of privatizing software, it becomes a means of keeping software free. The central idea of copyleft is that we give everyone permission to run the program, copy the program, modify the program, and distribute modified versions—but not permission to add restrictions of their own. For an effective copyleft, modified versions must also be free." – from The GNU Project, Richard Stallman, http://www.gnu.org/gnu/thegnuproject.html

[27] "An open source license is a copyright license for computer software that makes the source code available under terms that allow for modification and redistribution without having to pay the original author. Such licenses may have additional restrictions such as a requirement to preserve the name of the authors and the copyright statement within the code. One popular (and sometimes considered normative) set of open source software licenses are those approved by the Open Source Initiative (OSI) based on their Open Source Definition (OSD)." from http://en.wikipedia.org/wiki/Open-source_license

now secured through the Creative Commons Attribution - Share Alike[28]. The application of a standardized permissive copyright license guarantees the rights of users to copy, distribute, and modify the knowledge inside Wikipedia without asking permission of other individual users, or of the "leaders" of Wikipedia.

Wikipedia has an effective, if non-technical, "semantic" amount of modularization, which is derived from the very format of the encyclopedia. One can insert random *jibberish* into a Wikipedia page without damaging the rest of the language on the page, and an editor or any user can quickly remove the offending language, again without interfering with the ability of a reader to see and use "traditional" knowledge processing systems (e.g., the brain) to understand the content. The standard MediaWiki technical format renders content in a harmonious form, and embeds content versioning similar to software code versioning systems.

As institutions move towards a public role in distributed innovation, it will be vital to implement something akin to the technical separation of concerns into the governance of knowledge creation and distribution. It is a framework with a major role in existing distributed knowledge governance processes, and it carries deep implications for the academy: it promises to transform the role of the individual inside the academy by allowing more and greater access to knowledge, faster publishing and correction, more democratic peer review, at the same time that it may also allow less traditional actors to enter the knowledge governance systems, as editors, readers, critics, and hopefully at least occasionally, new partners.

However, the distributed systems of innovation we have been discussing disdain the larger institutions inside which the traditional actors of knowledge governance sit, and which play an enormous role in the exercise of intellectual property rights beyond the copyrighted creative work or software. The institution in this context is much more similar to a corporate firm than a collection of individuals, and thus needs its own theory of innovation, to evolve into something more modern and network-friendly. There is another new theory of knowledge governance, known as "Open Innovation," which is more closely associated with the firm as governor than the individual or the collective of individuals, which provides some ideas for the university.

---

[28] http://www.gnu.org/copyleft/fdl.html and http://creativecommons.org/licenses/by-sa/3.0/us/

The Open Innovation theory builds on the observation that an institution sits in an ecosystem of empowered individuals and other institutions, but that in a pre-network world the transaction costs of accessing the innovations of those actors was too high to justify. Thus, the institution develops its own internal knowledge creation and governance systems (technology transfer offices, tenure and review boards, etc). In a network culture there is the opportunity to connect more and more of those smart people to an institution's mission: to contribute to internal projects from the outside, to take a project that fails to gather internal support forward using outside funding, to generate novel projects outside and "spin into" new internal projects. All of this becomes possible as the transaction costs involved in the knowledge movement required drop.

Open innovation builds on the idea of knowledge "leakage"[29] from the firm or institution to the outside world, which is an informal kind of governance. This leakage comes from many sources depending on the kind of institutions. It could be publication, via scholarly journal, by patent, by conference presentation, informal conversation, or more currently via blog post or wiki edit. The core insight of Open Innovation is the ability to use the world outside an institution to generate internally useful knowledge - and the core dependency of open innovation in turn is the need to make the flow of knowledge in and out of an institution a purposeful thing, not a random process.

Another relevant element of Open Innovation theory is the importance of the business model. A world of purposeful information flow in and out of institutions is at odds with many of the business structures of the last 50 years - especially intellectual property rights. Copyrights govern

[29] "Recent scholarship has argued that the comparative success of the Silicon Valley high technology industrial district and failure of Route 128 outside of Boston, resulted from different patterns of inter-firm employee mobility which, in turn, led to differing patterns of industrial organization: network organization as opposed to traditional vertical integration. The cause of the different patterns of employee mobility is said to be cultural differences between California and Massachusetts. This paper offers a different causal analysis. After reviewing the new economic geography's emphasis on inter-firm knowledge transfers as an agglomeration economy, I focus on the critical role of employee mobility -- the vehicle for inter-firm knowledge transfers -- in facilitating second-stage agglomeration economies: those that allow the district to transcend its original product cycle and reinvent itself. "Gilson, Ronald J., The Legal Infrastructure of High Technology Industrial Districts: Silicon Valley, Route 128, and Covenants Not to Compete (August 1998). Stanford Law School, John M. Olin Program in Law and Economics, Working Paper No. 163. Available at SSRN: http://ssrn.com/abstract=124508 or doi:10.2139/ssrn.124508

the copying, distribution, and reuse of the documents containing actionable knowledge, from software to scholarship. Trade secrets and knowledge leakage on the public web are completely at odds with one another. And patents prevent institutions from acting on useful knowledge, even if the action would be far afield from the business concerns of the patent owner. Business models incorporate these knowledge "properties" as assets to be protected, and build infrastructures of lawyers and compliance offices precisely to prevent their flow out and usage in the external world. Thus, the business model often forms a block to the institution's adoption of an open innovation-based knowledge governance model, even if the ideas and theories of open innovation are attractive to the management and leaders of the institution.

Network-mediated knowledge governance in the university, then, sits inside a complicated context. Institutions must deal with changes in technology, pedagogy, incentive structures, libraries, intellectual property rights, business models, and more if they are to fully engage in either distributed innovation or open innovation models – or, ideally, both. It is clear that making open systems simple and scalable turns out to be a complex affair. And the standardization of the network is, so far, a technocratic one. It enables the emergence of the Open Innovation paradigm but does not yet enable the horizontal spread of Open Innovation, because each institution must develop its own bespoke strategy for knowledge governance (and rights to use the knowledge) needed for each situation.

In both the distributed and open cases for network-mediated knowledge governance briefly explored above, we saw the impact of interoperability as a baseline philosophy. Standard technical systems combine with standard knowledge formats to create open, generative solutions to problems. We also saw how the concept of interoperability applies from the pure technical "stacks" of standards to the legal realm via public licenses for knowledge products. We also saw that the existing paradigm cases of distribution and openness like Wikipedia and GNU / Linux exist outside traditional, "top down" approaches to knowledge governance - echoes of John Perry Barlow's famous declaration of independence of cyberspace[30] and David Clark's famous summation:

[30] "Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather." from http://homes.eff.org/~barlow/Declaration-Final.html

> "We reject: kings, presidents and voting.
> We believe in: rough consensus and running code."[31]

Yet interoperability, the basis for all of this emergent knowledge governance, is a choice. Like all choices, it can be undone, as Lessig has repeatedly warned[32]. Interoperability can decay through slow changes to technical standards, government regulation, and consumer choice - as in the emergence of "semi-closed" ecosystems of social networks like Facebook in which the interoperability is only enabled inside a walled garden[33]. The choice of licensing can also affect the downstream interoperability of a knowledge system, with even "open" licenses having negative impacts on knowledge integration and diffusion as unintended "environmental" consequences of an attempt to foster a sharing culture. [34]

2. Case Study: Genomics

Genomics is the study of the genomes of organisms. The field includes intensive efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping efforts. The field also includes studies of intragenomic phenomena and other interactions between loci and alleles within the genome[35] and the techniques of sequencing, genome mapping, data storage, and bioinformatic analyses. The wide range of genomics outputs means that the entire range of intellectual property rights come into play, and its relative youth as a field provides an ideal case study for knowledge governance systems in development and transition.

---

[31] A Cloudy Crystal Ball -- Visions of the Future" (PDF). 1992-07-16. http://xys.ccert.edu.cn/reference/future_ietf_92.pdf. Retrieved 2008-06-04. (Presentation given at the 24th Internet Engineering Task Force.)

[32] In particular, Code and Other Laws Of Cyberspace, 1998.

[33] The Generative Internet, Jonathan L. Zittrain, 119 Harv. L. Rev. 1974 (2006)

[34] "…a possible unintended consequence in the long run is the possibility of creating systems of data sharing that have embedded within them the seeds of license incompatibility—seeds that once planted may mature into future interoperability problems that were not only unanticipated and unintended, but that are also too complex to solve or even to understand completely. These problems may undermine the very goals of public data sharing and artificially limit data exchange and collaboration between communities—not for justifiable technical or scientific constraints but merely for license incompatibility reasons. There is evidence that such problems have already arisen throughout the open data community and may become more severe unless we take steps to address them at an early stage through developing a workable policy consensus." Science Commons "Comments on the Open Database License Proposed by Open Data Commons", http://sciencecommons.org/resources/readingroom/comments-on-odbl/

[35] http://en.wikipedia.org/wiki/Genomics

The Human Genome Project (HGP), a massive intergovernmental effort to sequence the human genome, can be studied through the lens of three knowledge products – narratives, data, and tools – that are at the core of knowledge governance in both industrial and academic genomics. This analysis will focus on the IP rights associated with the knowledge products, provide a short overview of the HGP, examine the changes in knowledge governance that developed over the course of the HGP, and close with the trends currently in play in post-HGP genomics.

2.A. Intellectual Property Rights and Knowledge Products

Copyrights, which govern the right of users to make and distribute copies of creative works, attach to narratives like journal articles, conference proceedings, posters, blogs, wikis, and other narrative forms of communication, as well as to bioinformatics and cheminformatics software. Copyright protection is typically transferred from the author to the publisher of information in communication systems that pre-date the Internet (especially journal articles) and the copyright is used by the publisher to extract revenues from subscribers. For software, there is a broad range of licensing options from classic open source models (more and more prevalent in bioinformatics) to very closed models, heavily protected by copyright licenses and charging high fees per seat.

Copyrights also factor into data and database protection: although raw "facts" like gene sequences are typically held to be non-creative works and thus not subject to copyright, varying levels of protection are allowed on the collective database itself depending on the national location of the database. U.S. law allows a very limited copyright on the elements of the database that represent "selection and arrangement"[36], while the U.K. and some other countries allow a stronger "sweat of the brow" copyright over a database that rewards the act of collection itself. In the European Union, a law was written specifically for database protection that creates a "sui generis" right akin to copyright, which governs databases whose servers are located in the EU[37].

---

[36] More detailed explanation on this issue see Carroll, Michael (2009) Copyright in Databases, at http://carrollogos.blogspot.com/2009/02/copyright-in-databases.html

[37] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML

Patents, which govern the rights to exclude others from and practice an invention, attach to innovations like isolated gene sequences[38], genetically modified sequences, disease mechanisms of action, engineered tools like stem cells, mice[39], drug compounds, among others. Patents are considered to be a critical element of the business of genomics and proteomics, and can bring great wealth to owners if the patent turns out to be a critical element of a drug or diagnostic that is taken to market.

Trade secret is also a powerful force in genomics and proteomics. Before publication, most academic research is held in secret, as is most corporate research before a project either leads to a patent or to abandonment.

In the context of knowledge governance, the natural contours defined by the use of intellectual property give further focus to our study. Those contours create a path to understand genomics and proteomics more generally, and also cast light on the broader biotech field itself. From pricing pressures in scholarly publishing[40], to the impact of a natural public domain data status[41], to the impact of community norms[42], legislative and judicial changes on research tools[43], our

---

[38] The possibility of patent of isolated genetic sequences is currently under debate in the American Courts in the case ACLU v. Myriad. ACLU is seeking invalidity on claims affirming that isolated genetic sequences are unpatentable subject matter. See more at http://www.genengnews.com/news/bnitem.aspx?name=54504126&source=genwire. Nowadays, there are more than 50000 USA patents in genetic sequences. See: http://dnapatents.georgetown.edu/

[39] The case of access and use of mice and genetic modified mice as research tools is explored in depth by Fiona Murray, http://fmurray.scripts.mit.edu/Published_and_Forthcoming_Papers.html

[40] The term serials crisis has become a common shorthand to describe the chronic subscription cost increases of many scholarly journals.[1] The prices of these institutional or library subscriptions have been rising much faster than the Consumer Price Index for several decades,[2] while the funds available to the libraries have remained static or have declined in real terms. As a result, academic and research libraries have regularly canceled serial subscriptions to accommodate price increases of the remaining current subscriptions. From http://en.wikipedia.org/wiki/Serials_crisis

[41] “Information that is created by or for the US government on this site is within the public domain. Public domain information on the National Library of Medicine (NLM) Web pages may be freely distributed and copied. However, it is requested that in any subsequent use of this work, NLM be given appropriate acknowledgment.” - http://www.ncbi.nlm.nih.gov/About/disclaimer.html

[42] The "Bermuda Rules" may sound like standards for lawn tennis, but in fact they are guidelines for releasing human sequence data. Established in February 1996 at a Bermuda meeting of heads of the biggest labs in the publicly funded genome project, the rules instruct competitors in this cutthroat field to give away the fruits of their research for free.” From Bermuda Rules: Community Spirit, With Teeth. Eliot Marshall, Science 16 February 2001: Vol. 291. no. 5507, p. 1192,

[43] “Fundamentally, Bayh-Dole shifted the incentive structure that governed the research and development path of federally funded tools by allowing institutions to own tools resulting from federally sponsored research and to exclusively license those tools.” Bayh-Dole: if we knew then what we know now. Sara Boettiger & Alan B Bennett, Nature Biotechnology 24, 320 - 323 (2006)

study of tools, narratives, and data provides a lens onto cooperation across most of the key actors and segments of the industry.

The role of copyright on papers fits into a complex ecosystem of knowledge governance, which ties together many of the stakeholders in more "traditional" scientific knowledge transfer. "Tools" are a key knowledge-embedded product in genomics, whether or not the inventor seeks a patent, and withholding of unpatented tools plays a critical role in the informal knowledge governance of the sector[44]. The relative dis-harmony of database copyright treatment will be examined later in the case study as it affects knowledge governance, but so far has been avoided through the aggressive use of the public domain as a legal treatment for data.

2.B. Human Genome Project and Data Governance

Completed in 2003, the Human Genome Project (HGP) was a 13-year, $3,000,000,000 project coordinated by the U.S. Department of Energy and the National Institutes of Health. During the early years of the HGP, the Wellcome Trust (U.K.) became a major partner; additional contributions came from Japan, France, Germany, China, and others. Project goals were to 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, store this information in databases, improve tools for data analysis, transfer related technologies to the private sector, and address the ethical, legal, and social issues (ELSI) that may arise from the project.[45]

While the entire project raised issues of knowledge governance, first we will examine the issues related to the datasets and databases created in the HGP, because the data governance regimes that emerged from the Human Genome Project served as the basis for setting norms for the development of common-based practices in the genomics field that last far beyond the HGP itself.

---

[44] "On the basis of survey responses from 507 academic biomedical researchers, we examine the impact of patents on access to the knowledge and material inputs that are used in subsequent research. We observe that access to knowledge inputs is largely unaffected by patents. Accessing other researchers' materials, such as cell lines, reagents, and antigens is, however, more problematic. The main factors associated with restricted access to materials include scientific competition, the cost of providing materials, a history of commercial activity on the part of the prospective supplier, and whether the material in question is itself a drug." Walsh, John P, Charlene Cho, and Wesley M Cohen. 2007 "Where Excludability Matters: Material v. Intellectual Property in Academic Biomedical Research"

[45] Human Genome Project home page, http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml

As a global distributed project, the HGP was forced early on to grapple with the issues of legal and technical interoperability, data acquisition and distribution, and scientific traditions of priority, publication, and citation. It was also a deeply asymmetric project - it was understood that a limited group of people could contribute since there was a lack of capacity and infrastructure. This limited funding to large sequence centers at major, well known, and powerful universities. Not many had the scientists, the labs or the machines to develop the study – a marked characteristic of differentiation when you compare the HGP with Open Source projects, where there is a democratization of means. Knowledge governance leading data *from the few to the many* would be required to facilitate an open system's later emergence.

However, after some time into the project development, the sponsors of the project – the governments and private funders – realized that the public data deposits were falling behind the rate of publicly funded data production. Worse, a private competitor (Celera) was rapidly accelerating the creation of a closed whole-genome sequence. The government and funder reaction was to send the key scientific leadership away to Bermuda (later Fort Lauderdale, FL) to work out the problem amongst themselves. [46] The basis was scientific, not legal, and deeply tied to the innate asymmetry of funding and the knowledge governance obligations it created:

> "if genome centres restrict their data and get preferential access to it, then some members of the community will no longer support monopolistic funding models (in which large centres sequence one genome after another without peer review of each project). Instead, they will demand the right to compete with these empires, especially for the most scientifically desirable genomes. Other scientists, especially bioinformaticians, will seek to relocate to the centres to gain the advantage of early data access. Data restrictions will therefore promote factionalization where we should be seeking efficiencies of scale, and centralization where we should be promoting diversity"[47].

---

[46] A detailed analysis of this is well beyond the scope of this paper, but Sir John Sulston's book The Common Thread contains a first-hand lively and knowledgeable account of the events.

[47] Salzberg, Steven, Ewan Birney, Sean Eddy, and Owen White, "Unrestricted Free Access Works and Must Continue," Nature 422 (2003).

The resulting 1996 agreement is widely known as the Bermuda Rules. This landmark agreement is not a legal construct, contract, license, or otherwise binding in a court of law. It simply represented the norms of the HGP sequencing community. And the rules are simple. First, take care of the backlog by releasing, immediately, all DNA strands longer than 1000 units; second, all new data goes on the web and into the public domain within 24 hours of coming off the machine.

Within this open governance regime the sequencing centers developed a strong competitive streak, which drove more and more data into the public domain, faster. One key requirement in the success of the Bermuda Rules was, in exchange for access, the application of a scientific publication norm: the centers depositing retained certain rights of first publication. But this again was a norm, not a legal requirement. Violators were in the realm of scientific publication and community judgment, not the courts.

There was great expectation that the impact of the release of the genome data would be that genomics companies would dominate the new face of drug discovery and development[48], which faded as it became clear that data on its own was not sufficient to provide the knowledge required to understand diseases or discover drugs. The publication of the complete human genome in the public domain also had a significant governance impact on companies whose business was to use trade secrets to protect their data products, and companies such as Celera were unable to continue a business model based on expensive subscriptions for data that was available on the web in the public domain.[49]

---

[48] Derek Lowe, who was there, wrote the following on the genomics craze: "most of the stuff you heard, at least in press releases and the like, ran to "Genomatronic Corp. announces that it has now filed patent applications (a whopping load of patent applications) on another huge, important swath of the vital human genome (remember, there's only one!), and reminds the industry that its back walkway is open on Tuesdays and Thursdays for Big Pharma to come crawling up it". Over at Megapharm, Inc., their opposite number, the fear was quite real that the Genomatronics of the world actually were staking out all the deposits of gold, and that all the drug targets in the world were going to end up owned by someone else – like those other big drug companies that were daily announcing huge deals with Genomatronic et al….It was easy for panic to set in. How much of the genome could possibly be left by now? We'd better do a deal while there's something to buy! After all, when you got down to it, these folks were right – there's only one human genome, and we're only going to read it for the first time once, and all the drug targets that will ever exist are in there – right? So why would you sit there and watch the competition walk off with all the good stuff? Right?"
http://pipeline.corante.com/archives/2009/01/19/ten_years_after_the_genomics_frenzy.php

[49] Specifically within the USA, over the past ten years, companies like Celera, Incyte, Human Genome Sciences, Millenium, and more have exited the foundational data market, with Celera the most extreme example - abandoning the database market entirely by depositing their private genome sequence directly into Entrez . This is a reflection

Another key aspect of the HGP was the systematic investment in the technical infrastructure to distribute the sequences, a combined effort of the various governments involved that included nightly sequence harmonization across the various data repositories. The emergence of the U.S. National Center for Biotechnology Information (NCBI) was essential to the success of the project, demonstrating that technical accessibility is part of knowledge governance as well, and also developing some of the early integration of technology access with policy access – the NCBI not only clearly marks government data as public domain, but will not accept data whose depositors request controls based on intellectual property into its molecular databases.[50] Knowledge in this context must therefore also be studied with an eye towards technical architectures and their impact on governance.

The norms that emerged from the HGP were the inspiration for the norm-setting process in its successor, the "HapMap" project of human genetic variation,[51] and many of the same technical infrastructures were expanded to include the variation data alongside the genome. However, different from the origins of the HGP, when the HapMap was born, the Open Source Movement was a well developed and studied movement. The founders were inspired by Open Source to adopt an "open click-wrap" data license that tried to regulate publication processes and intervene in the exploration (more precisely – the abuse) of patents that may have emerged from the HapMap outputs[52].

This licensing approach was abandoned for an unregulated environment running under the Bermuda Rules after a few years of operations. The fear of patent enclosure had been

---

both of changing market conditions and of the growing economic power and value of the unregulated-open systems. Merck, one of the world's largest pharmaceutical companies, added to this growth by the deposition of the Merck Gene Index into the Entrez system, which was a strategy to establish pre-competitive gene sequences to avoid widespread gene patent "thickets" but had the secondary consequences of increasing market power in an unregulated-open space and creating market standardization around the Entrez technology platform as well as the public domain .

[50] "NCBI itself places no restrictions on the use or distribution of the data contained therein. Nor do we accept data when the submitter has requested restrictions on reuse or redistribution." - http://www.ncbi.nlm.nih.gov/About/disclaimer.html

[51] Whereas the Human Genome Project set out to map the DNA that is common to us all and makes us human, the HapMap project set out to map the individual genetic variations that make us individuals.

[52] More Than Two Years After Its Inception, Int l HapMap Project Defines Data-Release Policies, GenomeWeb, December 18, 2003. http://www.genomeweb.com/dxpgx/more-two-years-after-its-inception-int-l-hapmap-project-defines-data-release-pol

ameliorated by the dedication of so much data to the public domain, so the patent aspects of the contract were felt unnecessary, and the unintended *knowledge governance impact* of the contract was that the HapMap data could not be integrated into the HGP data without creating legal contamination. Thus, the contract was lifted and the norms moved into place instead[53].

The HGP is in many ways a paradigmatic case of the shift in knowledge governance from privacy and withholding to open sharing. The perfect mixture of policy, funding, norms, law, and technical infrastructure came together to open up the genome to all – when an alternate outcome could have easily occurred. The genome is now fundamentally open data, legally and technically[54]. Its entirety can be downloaded without registration, and redistributed. It can be annotated, visualized, and built upon. It leverages standard data formats, data repositories, software tools, and more. And we see the long-term environmental impact of "conserving" the genome in a realm where the knowledge governance of genome data was run by the scientists, not the courts, in the explosion of downstream knowledge products emerging from the HGP and its successors.

From the genome as a foundational base we see many distributed efforts emerging, as we might expect from an open source or wiki approach. Distributed annotation systems emerged to mark up the genes on the genome[55], and an entire new field - synthetic biology - erupted, using the information gleaned in the genome sequence to create standard biological "parts" to be used in biological programming systems[56]. Yet to achieve this, the HGP represented a years-long investment in fundamental data generation with an unclear outcome, and required the creation of significant funding streams to support technical distribution systems. Scientists had to come together and develop new norms and ethics for knowledge governance and distribution, not to mention standards for annotation and reuse as well as software tools and systems that made the genome useful, accessible and interoperable.

---

[53] Netwatch: HapMap Lifts Data Restrictions. *Science* 24 December 2004: Vol. 306. no. 5705, p. 2167

[54] There is a knowledge utilization issue related to gene patents that is both real and significant but sits outside the scope of this short case study. Readers wanting more information on this topic are encouraged to examine the excellent work of many scholars, in particular Robert Cook-Deegan's *The Gene Wars: Science, Politics, and the Human Genome.* New York and London: WW Norton & Co.; paperback 1996, 1994.

[55] Dowell RD, Jokerst RM, Day A, Eddy SR, Stein L (2001). "The distributed annotation system.". *BMC Bioinformatics* 2: 7

[56] Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 2000 Jan 20;403(6767):335-8.

2.c. Governance of Narratives and Tools

Although we have focused on data so far, we can also contemplate the outputs of the Human Genome Project as they formed narratives and tools. The HGP yielded thousands of these knowledge products and a concurrent explosion of startup companies in genomics and proteomics, which peaked in the genomics "bubble" of the late 1990s. After the end of the genomics bubble, genomics companies trended away from data and towards "platforms" typically based on tools: massive-scale sequencing, cloning expertise, and functional genomics.

What each of these companies has in common is a foundation in the academic literature, which is where the vast majority of theoretical research was published, and in academic laboratories, where many tools were first developed before being "polished" inside corporate structures for sale. Each of these classes of knowledge products underwent significant and sweeping changes over the 13-year history of the HGP. We will first examine the changes in governance over narratives as the publishing industry adapted to the Internet, and then the changes in invention governance in the face of a changing legislative and judicial context.

Over the same time period as the HGP, scientific scholarly publishers saw in the Internet an increase in their ability to leverage technology. Similar to the music and movie industries, a small but dominant group of publishers (e.g. Elsevier, Nature, Springer, Wiley-Blackwell, and others) explored models in which their traditional methods of selling physical copies of narrative scholarly content gave way to new business models of renting access to content via copyright licenses.

Unlike music, however, profits exploded in science and technology publishing. Combined with practices like "price bundling" in which university subscribers were forced to subscribe to less-popular journals in order to access the most desirable titles, and by a powerful reliance on copyrights enforcement, the overall cost of access to scientific narratives outpaced the cost of living increase by 600% over the course of the biotech industry, from beginning through bubble and into the present day[57]. In reaction to this pricing crisis, the Open Access[58] movement[59] evolved into a fully empowered active system in the narrative space, opening space for new open business models and for institutional mandates for "author self-archiving".

Production and distribution of narratives exists at a different cycle than data and tools, because the narrative endures two completely separate cycles of production: knowledge generation in the lab, and article production. The former is less our concern here but can take years, and if the experiment "fails" the narrative frequently never is produced into an article out of laboratory notebooks. The emergence of blogs and wikis may eventually have a powerful role in capturing these "failed" narratives, but as yet have had little impact.

The article production cycle is shrinking with the advent of the digital journal. Where it may once have taken a year or more to get an article from submission into a print media journal sent via mails, the cycles now can be as short as a few weeks in the case of PLoS One.[60] The entire production cycle is much more efficient, although journals with high rejection rates continue to demonstrate the longest production cycle. The defining factor in these long cycles is the peer review process, which can be aided by technology but remains inherently slow, as it is mediated by the social networks of a discipline. Distribution of the narrative has undergone similar transformations as other digital media, with the ability to email PDFs of papers, post copies of

---

[57] American Library Association, Economics of Scholarly Publishing, http://www.ala.org/ala/mgrps/divs/acrl/issues/scholarlycomm/scholarlycommunicationtoolkit/librarians/librarianeconomics.cfm

[58] "By "open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited. Budapest Open Access Initiative, http://www.soros.org/openaccess/read.shtml

[59] Open Access Overview, http://www.earlham.edu/~peters/fos/overview.htm and "The Open Access Movement: opportunities and challenges for developing countries" - Diplo Foundation, Internet Governance Program – 2007. (this article was accepted at "The Politics of Intellectual Property," at the ECPR Joint Sessions, Rennes, France, April 2008 and First Pan-African Forum on Open Education Resources (OER), Ghana, Africa, May, 2008). Brazilian Law Journal: Revista Criação, N. 1, Vol. 1, 2008, IBPI - Brazilian Institute of Intellectual Property. @http://campus.diplomacy.edu/env/scripts/Pool/GetBin.asp?IDPool=3737

[60] PLoS ONE is an open access, "online only", scientific journal from the Public Library of Science. It covers primary research from any discipline within science and medicine. All submissions go through a rigorous, internal and external pre-publication peer review but are not excluded on the basis of lack of perceived importance or adherence to a scientific field. The PLoS ONE online platform has post-publication user discussion and rating features. http://www.plosone.org/home.action

pre-print articles, and hyperlink articles into websites, emails and more. The ability of individuals to circulate their work marks a major change in knowledge distribution. Thus, we see echoed in the narrative governance the importance of the technologies as a dimension of knowledge governance.

A key inflection point towards openness in narrative governance came when the USA government flexed its market power as a funder of research, and implemented a mandate that all taxpayer-funded health research should be available under Open Access terms[61]. Deposits of scientific research skyrocketed in response.[62] Another key point in the Open Access market came in late 2008, when Springer purchased the Open Access publisher BioMed Central, whose revenues approached $15M per year.[63] Both of these occurred against, again, a technical backdrop in which governments created technical repositories to host open narrative content, just as with open data for the HGP.

The governance of tools in genomics underwent similar transformation over the life of the HGP. Tools were also dramatically affected by changes in knowledge governance via changes in the IP environment. Court cases allowing the patenting of modified genes[64], and legislation encouraging universities to acquire and license patents on government-funded research[65], paralleled the genomics bubble with a dramatic increase in the number of patents filed by universities on biology.[66]

---

[61] "The NIH Public Access Policy ensures that the public has access to the published results of NIH funded research. It requires scientists to submit final peer-reviewed journal manuscripts that arise from NIH funds to the digital archive PubMed Central upon acceptance for publication. To help advance science and improve human health, the Policy requires that these papers are accessible to the public on PubMed Central no later than 12 months after publication." NIH Public Access home page. http://publicaccess.nih.gov/

[62] NIMHS statistics are representative: http://www.nihms.nih.gov/stats/

[64.] http://www.ncbi.nlm.nih.gov/sites/gquery

[63] Open access publisher BioMed Central sold to Springer. Ivan Oransky, Scientific American, Oct 7, 2008

[64] Diamond v. Chakrabarty, 447 U.S. 303 (1980), was a United States Supreme Court case dealing with whether genetically modified micro-organisms can be patented.

[65] The Bayh-Dole Act or University and Small Business Patent Procedures Act is United States legislation dealing with intellectual property arising from federal government-funded research. Adopted in 1980, Bayh-Dole is codified in 35 U.S.C. § 200-212[1], and implemented by 37 C.F.R. 401[2]. Among other things, it gave US universities, small businesses and non-profits intellectual property control of their tools and other intellectual property that resulted from such funding. http://en.wikipedia.org/wiki/Bayh–Dole_Act

[66] K. Jensen and F. Murray (2005). "Intellectual Property Landscape of the Human Genome." Science, Vol. 310, 14 October 2005, pp. 239-240.

Biological tools can range from simple biological materials that can be generated from standard protocols (similar to recipes in the kitchen) and everyday lab materials to complex living systems like genetically modified mice. Some biological materials are breakthrough products on their own, like the first human stem cell lines, and can take years to develop in the first instance, though their living nature allows them to be "cultured" and grown again and again after the first successful cycle.

The importance of biological materials has grown over the last fifteen years, as research tools like genetically modified mice and stem cells became critical to replicating published research, but access to those materials was (and is) frequently blocked by patent rights and by competitive withholding by scientists, research institutions, companies, and other stakeholders.[67] Strikingly, policies that echo the data and narrative deposit requirements have been slow to emerge and difficult to implement – a 1998 working group of the US National Institutes of Health recommended open and pre-competitive access to research tools generated with taxpayer funds, but that recommendation has never been associated with mandates, metrics, or other systems that would encourage a transition to a true governance policy for tools (whether patented or not)[68].

While there is not a centralized repository for tools and tools like the NCBI for data and narratives, the market for tools is also affected by the establishment of biological resource centers (BRC) into which funders sometimes mandate deposit of knowledge products developed under funded research. Tool distribution tends to follow publication (and depending on the perceived economic value of the tool, patent applications) and again differs depending on the kind of tool, existence of BRCs in that class of tool facilitating distribution on behalf of the

[67] The market for research tools is hard to characterize in a general fashion and tends to be related more to the individual tool, like a method for generating human stem cells, than to a broader perspective economically. Plasmids tend to be traded without recourse to economics or patents, while genetically engineered mice almost always include patent licenses.

[68] Report of the National Institutes of Health (NIH) Working Group on Research Tools, Presented to the Advisory Committee to the Director, June 4, 1998

scientist, and popularity of the material.[69] Distributing and governing tools is inherently more expensive than data and narratives as the tools are non-digital assets.

2.D – Knowledge Governance Trends

Policy trends in genomic and proteomic research unambiguously favor openness and unregulated access. The most notable policy trend in publishing narratives for genomics and proteomics is the powerful shift towards open access to research articles symbolized by the public access policies implemented the US Government (NIH most obviously) and endorsed by governments and institutions across Europe, in Australia, and elsewhere[70]. The original policy direction of a few funders is now echoed by private foundations[71] and research universities[72].

A similar set of trends is also evident in foundational data, which tends to be created as a "big science" collaborative project, and in which the funds to disseminate data are part of the funding contract. The success of the HGP and related projects have created an expectation of data sharing of what we might call "foundational" information to serve as a basis of future experimentation.[73] At the smaller scale, the experimental data generated in laboratories is subject to less formal

[69] Other tools in this context are laboratory robots, which are essential to the data production cycle. Such robots include microarrays for rapid analysis of gene expression, high-throughput genetic sequencers, flow-assisted cell sorters, and more. These tools are commodities available from catalogues at varying price levels (even on eBay) or as services from genomics core facilities to produce data on-demand at levels previously unthinkable. The only restrictions on data production capacity where these machines exist is funding and ability to utilize the data. Distribution of the generated data tends to depend on technical capacity and class of data - if the NCBI system accepts deposit of a class of data, distribution of that data tends to be dominated by NCBI. Otherwise, there is a wide range of knowledge governance behavior ranging from posting of data files on laboratory websites to "email me to ask for the data" to outright withholding.

[70] http://www.earlham.edu/~peters/fos/newsletter/01-02-10.htm#2009

[71] "Autism Speaks, the nation's largest autism advocacy organization, announced that effective December 3, 2008, all researchers who receive an Autism Speaks grant will be required to deposit any resulting peer-reviewed research papers in the PubMed Central online archive, which will make the articles available to the public within 12 months of journal publication. This new policy will make the results of Autism Speaks-funded research easily accessible - at no charge - to individuals with autism, families and other advocates, as well as interested researchers. Autism Speaks is the first U.S.-based non-profit advocacy organization to institute this public access requirement." http://www.earlham.edu/~peters/fos/2008/11/autism-speaks-adopts-oa-mandate.html

[72] The Registry of Open Access Repositories lists 204 institutional policies on open access at the time of writing. http://www.eprints.org/openaccess/policysignup/

[73] By foundational, we mean data that is standardized and common – the genome would be an example, as would be the height of Mount Everest – that takes an enormous amount of funding to create, and can serve as the foundation for much other research. This is a term of the authors' invention, to distinguish from the kinds of data generated in a laboratory by the perturbation of a system.

policy requirements though the expectations of scientist and funder, informal and formal, are tilted towards sharing, with NIH grant proposals requiring the submission of data sharing plans. However, optimism for open data must be tempered with the reality of data sharing, which is difficult, expensive, and often unsatisfying when the repository is not the NCBI.[74]

Research tools like stem cells and plasmids are subject to completely voluntary sharing policies under the NIH system and though some private research foundations do both mandate and fund the sharing of tools, this is the exception rather than the rule. The counterbalancing trend is brought by the continuing reliance on innovation systems legislation as an influence for universities to patent research tools. Innovation legislation (such as the Bayh-Dole Act in the USA) is also occasionally brought into play in the sharing of data from which patentable tools might be drawn. However, the lack of standardization in contracts and technology means that very high percentages of tool sharing requests result in denial[75].

Another market aspect of genomics and proteomics is the market for research funds, which is dominated by the pursuit of NIH "R01" grants, the oldest and most common type of government funded research in the world of biotech within USA. In addition to R01 funding, researchers also compete for funds from private research foundations (frequently focused on specific diseases, but including some large research foundations like the Howard Hughes Medical Institute and the Wellcome Trust). Competition for grants is intense, and leads to some instincts towards data and tools withholding, while placing an enormous pressure on the researcher to publish narrative in "high quality" journals.

[74] "Molecular biology data are subject to terms of use that vary widely between databases and curating institutions. … While a few public domain policies are standardized, most terms of use are not harmonized, difficult to understand and impose controls that prevent others from effectively reusing data. Identifying a small number of restrictions allows one to quickly appreciate which databases are open. A checklist for data openness is proposed in order to assist database curators who wish to make their data more open to make sure they do so." Check Your Data Freedom: A Taxonomy to Assess Life Science Database Openness, Melanie Dulong de Rosnay

[75] "In the university – to – university setting, estimates range from delays of over a month for between 11% and 16% of requests, "a substantial delay in a fast-moving field," to estimates that there are routine delays of over 6 months for 20% of requests and over 2 months for 42% of requests. Studies also show increasing rates of outright denial of requests, and of abandonment of "promising research projects" because materials are not received. In the commercial-university arena, with no standardized agreements at all, most observers believe the situation is worse. Commercial-academic denial rates are estimated to be nearly twice those in the academic-academic context (33% to 18%)." Empirical Data About Materials Transfer Problems, Science Commons, http://sciencecommons.org/projects/licensing/empirical-data-about-materials-transfer/

2.E. Genomics: Early Conclusions

From this necessarily brief study, we can draw some conclusions about knowledge governance and its changes in reaction to the emergence of the technical revolutions of biotechnology / digital communications, and the legal-social revolutions of the distributed creation of knowledge.

First, each knowledge product carries its own set of informal and formal governance regimes that affect distribution and transfer. Intellectual property rights are one of the factors here, but competitive withholding (either for academic advancement or corporate value) has a major impact as well. Second, the infrastructure and incentives for distribution and re-use are complex yet essential. Data is of little value if there is no infrastructure to make it comprehensible, and tools represent a valuable commodity for hoarding under the current knowledge governance system. And third, in order to achieve real changes in knowledge governance regimes, a coalition of funders, research institutions, knowledge publishers, governments, and technical infrastructure providers seems essential.

3. National Innovation Policy

The three conclusions noted above hint at the direct role in the stimulation of innovation that knowledge governance must play. But although innovation policy has become a core element of policy-making in both the developed and developing world as countries engage in the information society process and compete in the globalized world, there is little "network centric" knowledge governance included as part of the process. Instead – looking to the example of the United States and later Asian countries and as a answer to the OECD's call related to the technology transfer of public funded research – countries of the North and South have implemented national legislation and public policies based in large part on the expansion and management of intellectual property. These trends are specifically intended to emulate the success of the United States in innovation, in a traditional innovation context, but may have significant and potentially toxic effects on emerging, distributed or open knowledge governance regimes.

Kenneth Shadlen argues that "National patent regimes influence trajectories of industrial development and governments' capacities to address humanitarian concerns. As pillars of national systems of innovation, patent regimes drive technological change and shape trajectories of knowledge-creation and knowledge-diffusion."[76] The "traditional innovation belief is that Intellectual Property should serve as the primary measurement of innovation, as a proxy for knowledge creation and governance, and also act as the primary safe-guard to attract investment in research and development, and not just as a promoter of diffusion[77].

The policies around traditional innovation focus on institutions, mainly universities and companies, as the key bridge between design and manufacturing. This focus reflects the experiences of the American Bayh-Dole legislation and implementation. Countries as diverse as Brazil, China, South Korea, India, Japan, France and Finland have implemented variations on the Bayh-Dole legislation into their innovation laws and policy arrangements.[78]

The USA's Patent and Trademark Law Amendment Act 96-517 of l980, the formal name of the Bayh-Dole Act, is the legal framework for transfer of university generated, federally funded inventions to the commercial market place. Bayh-Dole "swept away the patchwork of individual agency-controlled IPAs and instituted a uniform federal patent policy[79] for universities and small

[76] Ken Shadlen. Revolution, Reform and reinforcement: Latin America Responses to the Globalization of Intellectual Property. Working paper presented at the ECPR 2008. (pre-print version with the author)

[77] Research and development is considered the fundamental step to what economists call innovation of product and process to market, while diffusion is just related to the incorporation of existing technology to satisfy necessities of cost reduction, standardization of products and process and manufacture efficiency.

[78] Echoes of Bayh-Dole? A survey of IP and technology transfer policies in emerging and developing economies. In: Krattiger A, Mahoney RT, Nelsen L, Thompson JA, Bennett AB, et al., editors. Intellectual property management in health and agricultural innovation: A handbook of best practices. Oxford (UK): MIHR, and Davis(CA): PIPRA. pp. 169–195.

[79] This opinion is not uniform among authors that discuss the Bayh-Dole effects, since the Bayh-Dole, as determined in its Section 202 was first intended to regulate universities, small business and non-profits, as will be seen further in the paper. Jordan J. Baruch, Assistant Attorney General, during its testimonies, as quoted by Eisenberg, affirmed that: "The Bayh-Dole bill was only a partial solution that did nothing to unify patent policy across agencies; indeed, if the right of large contractors were still to be governed by the inconsistent practices and policies of the various agencies, a new set of statutory rules applicable only to nonprofit institutions and small business would merely add a twenty-seventh policy to the twenty-six inconsistent sets of rules and regulations already an the books." Supra note 10, at 1694. The concern was to avoid market power concentration and promote widespread commercialization of inventions and the flourishment of new companies (start ups), since large business are generally pictured as "short-sighted, risk-averse, and predatory, more likely to suppress new technologies than to adopt them" (Eisenberg, Supra note 10, at 1696). This problem finished in 1984, when an amendment, endorsing President's Regan memorandum previously sent to the head of the agencies, extended the authorization of title retention to large business.

businesses under which they obtained the rights to any patents resulting from grants or contracts funded by any federal agency."[80]

Under this 1980 law, as amended, the title to inventions made with government support is provided to the contractor if that contractor is a national small business (and after 1984 also large business), university, or other non-profit institution such as hospital and research institutions[81]. The legislation is intended to use patent ownership as an incentive for private sector development and commercialization of federally funded research and development (R&D).

The influence of Bayh-Dole is so great that it is worth examining its terms closely. In its section 200, the Act establishes as objectives for the use of the patent system to: (a) promote the utilization of inventions arising from federally supported research or development; (b) encourage maximum participation of small business firms in federally supported research and development efforts; (c) promote collaboration between commercial concerns and nonprofit organizations, including universities; (d) ensure that inventions made by nonprofit organizations and small business firms are used in a manner to promote free competition and enterprise without unduly encumbering future research and discovery; (e) promote the commercialization and public availability of inventions made in the United States by United States industry and labor; (f) ensure that the Government obtains sufficient rights in federally supported inventions to meet the needs of the Government and protect the public against nonuse or unreasonable use of inventions; and (h) minimize the costs of administering policies in this area.

[80] Supra note 6, at 778.

[81] 35. U.S.C. Part II. Ch. 18:202: "Disposition of rights: (a) Each nonprofit organization or small business firm may, within a reasonable time after disclosure as required by paragraph (c)(1) of this section, elect to retain title to any subject invention." The Section follows to established exceptions to the rule by saying: "Provided, however, that a funding agreement may provide otherwise (i) when the contractor is not located in the United States or does not have a place of business located in the United States or is subject to the control of a foreign government, (ii) in exceptional circumstances when it is determined by the agency that restriction or elimination of the right to retain title to any subject invention will better promote the policy and objectives of this chapter (iii) when it is determined by a Government authority which is authorized by statute or Executive order to conduct foreign intelligence or counter-intelligence activities that the restriction or elimination of the right to retain title to any subject invention is necessary to protect the security of such activities or, (iv) when the funding agreement includes the operation of a Government-owned, contractor-operated facility of the Department of Energy primarily dedicated to that Department's naval nuclear propulsion or weapons related programs and all funding agreement limitations under this subparagraph on the contractor's right to elect title to a subject invention are limited to inventions occurring under the above two programs of the Department of Energy. (…)."

With the passage of the Bayh-Dole Act, universities began to develop and strengthen their internal expertise to engage in the patenting and licensing of inventions, through the establishment of entirely new technology transfer offices, building teams with legal, business, and scientific backgrounds. The Association of University Technology Managers (AUTM) shows that its membership increased from 113 members in 1979, to 691 in 1989, 2178 in 1999. The table below from the 2006 AUTM Licensing Survey shows how many Universities per year, since the 70s, started their Technology Transfer Program, which usually, implies the organization of a Transfer Technology Office (TTO). AUTM's report also shows that the TTO's staff levels are steadily increasing[82]. Creating and commercializing intellectual property have become one of the most important institutional objectives in various academic settings, a change usually credited to Bayh-Dole and traditional innovation.

It is important to juxtapose the centrality in traditional innovation of the institution as a closed entity, gaining and licensing intellectual property, against the role of the institution (company or university) in the Chesbrough open innovation paradigm[83], as well as against the role of the university as the host of a set of individuals participating in distributed innovation systems across traditional academic lines. The national innovation systems are set up to favor the traditional role, bolstered by academic incentives, high transaction costs for change, and general inertia towards organizational innovation.

In the traditional innovation context, an important role is played by the Research Universities that are positioned in a relational context with the government and the industry. This context is explained as a process in which the University adds industry-like responsibilities to its core mission of generation and dissemination of knowledge through education and scholarly research, in order to engage with the government and industry in the promotion of economic development.[84] Inside this relation, the inherent tension of the academia-industry relationship,

---

[82] See in general: AUTM. FY 2006 AUTM Licensing Survey. (2006) Available from: http://www.autm.net/events/file/AUTM_06_US%20LSS_FNL.pdf

[83] "Open innovation is the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation, respectively. [This paradigm] assumes that firms can and should use external ideas as well as internal ideas, and internal and external paths to market, as they look to advance their technology." Henry Chesbrough, *Open Innovation: Researching a New Paradigm* http://www.openinnovation.net/

[84] Specifically, the Sabato Triangle and the Triple Helix (in its I, II and III versions). See Failure and success: the fate of industrial policy in Latin America and South East Asia. Henry Etzkowitz and Sandra N. Brisolla, Research Policy, Volume 28, Issue 4, April 1999, Pages 337-350 and Etzkowitz's The Triple Helix: University-Industry-Government

and in particular, the role of patenting, has been the focus of significant analysis and discussion, particularly within the academic community[85].

Facing the challenges of national development and the goal of independent participation in the global economy, Brazil, for instance, developed a national innovation system (NIS) internalizing these innovation theories. Its NIS emulates foreign laws, such as the North-American Bayh-Dole University and Small Business Patent Act of 1980, as well as the French Loi Sur L'innovation et la Recherche from 1999. The main results were the NIS from 2003 and the Innovation Law from 2004 – among other norms related to tax incentives – both finalized and structured under the Luis Inacio Lula da Silva ("Lula") administration. Under the NIS, Brazil created priority areas such as biotechnology, pharmaceuticals (medicines and vaccines), biomass, information technology and software among others and developed the fundos setoriais ("sectoral funds") under the Ministry of Science and Technology budget to foster the expansion of research and its translation into commercially valuable goods and services. There is particular privilege for private-public partnerships, which receive tax incentives and other benefits. A parallel was recently developed under the coordination of the Development, Commerce and International Trade Ministry, called the Productive Development Plan (PDP) of 2008 that is not yet fully coordinated with the NIS under the work of interministerial commissions. All of these laws recognize the interconnections and interdependence between the knowledge institutions of industry, academia, and government.

However, none of these industrial policies have taken into consideration non-traditional knowledge governance arrangements like the open and distributed ones discussed in section 1. In the rush to emulate Bayh-Dole, many key elements of emerging, network-centric innovation policy, including new approaches to the patent system, remain unexamined by Brazilian policymakers.

Innovation in Action.

[85] See generally Eisenberg, Public Research and Private Development: patents and technology transfer in Government-Sponsored Research. 82 Virginia Law Review 1996; see generally Marcia Angell, M.D., Is Academic Medicine for Sale?, 342 NEJM 20 (May 18, 2000); see generally Arti Kaur Rai, Regulating Scientific Research: Intellectual Property Rights and the Norms of Science, 94 Nw. U. L. Rev. 77 (1999); see generally F. Scott Kieff, Facilitating Scientific Research: Intellectual Property Rights and the Norms of Science – A Response to Rai and Eisenberg, 95 Nw. U. L. Rev. 691 (2001).

Bayh-Dole was designed almost 30 years ago, before the network transformed our daily life. Bayh-Dole systems assume significant social, financial, legal, and institutional infrastructure to take innovations out of the academic/non-profit world to manufacture and marketing. Such legislation has the potential to significantly unbalance intellectual property regimes in countries that implement a variation of it. Bayh-Dole will require localization and adjustment to local knowledge governance realities if it is to work in Brazil as it has arguably worked in the USA[86]. Bayh-Dole "porting" is even more complex when one notes that developing nations in general do not present the cultural knowledge exploitation ecosystem inside which the Bayh-Dole functions in the USA, such as a strong private equity market, technology transfer community, or an entrepreneurial support system.

There is also a significant set of interlocking factors that can affect the effectiveness of Bayh-Dole-like legislation in a localizing country. Bilateral agreements and national legislation can inhibit the effectiveness of innovation policy by creating strong incentives for south>north innovation expropriation (especially if there is no local infrastructure to exploit local innovations). Other factors can retard innovation by making local forms of cultural or scientific business innovations difficult via expansive use of contract restrictions, anti-circumvention TPM (technological protection measures) laws, or expropriation of traditional knowledge.

The University has a real role to play here, as a "public space" (Piore and Lester 2004). The University represents a well-placed actor to foster and host arrangements of networked knowledge creation and distribution, adding to the Triangle or Helix of University-Industry-Government a fourth and empowered actor, the networked universe of individual user-innovators. Through the application of "commons" techniques like open licenses and through the

[86] "The Bayh-Dole Act has been seen as particularly successful in meeting its objectives. However, while the legislation provides a general framework to promote expanded utilization of the results of federally funded research and development, questions are being raised as to the adequacy of current arrangements. Most agree that closer cooperation among industry, government, and academia can augment funding sources (both in the private and public sectors), increase technology transfer, stimulate more innovation (beyond invention), lead to new products and processes, and expand markets. However, others point out that collaboration may provide an increased opportunity for conflict of interest, redirection of research, less openness in sharing of scientific discovery, and a greater emphasis on applied rather than basic research. Additional concerns have been expressed, particularly in relation to the pharmaceutical and biotechnology industries, that the government and the public are not receiving benefits commensurate with the federal contribution to the initial research and development." Wendy H. Schacht, The Bayh-Dole Act: Selected Issues in Patent Policy and the Commercialization of Technology Congressional research Services. (2006 - Updated October 5, 2007), at Summary. Available from: http://italy.usembassy.gov/pdf/other/RL32076.pdf.

use of cyberinfrastructure, the University can create a public space not just for its internal clients (student-faculty) but also public spaces in which the collected set of users on the network can connect through open educational resources and materials to the formerly closed Industry-University-Government innovation system.

For this, the University will have to develop a sufficiently complex internal policy of knowledge governance to allow the open and distributed knowledge models to emerge, concepts that bring together some important themes such as user reinvention, the economics of open source, open licensing, technologies of cooperation and collective action, and cyber-infrastructure for national and international collaboration and access to and transfer of knowledge. Also, the University will need interlocking governmental policies that support infrastructure development for non-traditional innovation and a legislative and judicial compromise to bar the expansion of the Intellectual Property rights, while Industry will need to develop business models that can exploit innovation of every form, including those based on non-traditional approaches to Intellectual Property.

Conclusions

The world of knowledge creation, distribution and governance sits at a crossroads, where the potential of new network driven systems hits the reality of traditional institutionally mediated knowledge governance. This is particularly the case in areas, such as science, that require more infrastructure and tools than are available from the consumer and commercial internet. Infrastructure brings institutions to the table, and institutions bring many traditional roles of knowledge governance into the conversation, which interact uneasily at best with the network systems and cultures.

There are few, if any, “easy answers” to the questions of what an institution, a government, a funder, or an individual should do in the short term. But there are early lessons to draw from the successes in knowledge access, formation and distribution outside the traditional system that provide clues.

First, the principle of interoperability should not be confined to technology, but instead should inform decisions ranging from policy to intellectual property to institutional arrangements and forward. Legal implications can reach through software to touch technology, content, knowledge products, and more - and interoperability as a design principle represents “good taste” in knowledge governance, as it both empowers those with the current capacity to participate in innovation and those who have not yet acquired that capacity. Practicing this principle of interoperability and separation of concerns means that we do not merely create infrastructure that serves today’s knowledge problem, but that can be extended and built upon to serve many knowledge problems in the future, most of which we cannot yet see. The leading expression to be here followed is *design it so it can attain interoperability*.

Second, the role of democratized access to infrastructure is essential. Open systems, be they legal, technical, or policy, that are not designed with interoperable infrastructure in mind are likely to yield unintended consequences. We saw this in the case study with the HapMap license, which blocked database integration as a side-effect of attempting to enforce patent openness. By violating the separation of concerns, an open knowledge governance attempt in patents resulted in a non-interoperable governance reality in data. Thus, policy and governance should not only tend towards open infrastructure, but contemplate the “environmental impact” of specific decisions on the availability of infrastructure.