**Cooperation Project**

**Progress Report for Ford Foundation**
**June/2009**

*Berkman Center for Internet and Society*
*Harvard University*
**Principal Investigator: Professor Yochai Benkler**
**Research fellows: Carolina Rossini**
**(ICP) and Aaron Shaw (OCR)**

**EXECUTIVE SUMMARY**

**Commons-Based, Cooperative, and Peer Production as Strategies for Development**

*Progress Report, Ford Foundation, June 19, 2009.*

Despite the growing popular and academic recognition of the importance of commons-based, cooperative, and peer production, there is still relatively little sustained academic work that studies both the scope and micro-foundations of these phenomena. Together, the rise of commons-based collaboration and production encompasses a class of innovative and creative practices whose outputs could be freely available to support human development in a global, networked information economy and society.

The purpose of the grant was to allow the Berkman Center to develop a more comprehensive map of current practices, and a basic set of methodological tools and approaches, to allow the continued study of commons-based practices, both online and offline, as well as large-scale networked cooperation. To this end, the proposal envisioned two tracks within a single study.

One track would focus on defining verticals within existing industries, and would map those industries in terms of the degree to which open and commons-based practices were used as compared to proprietary approaches. This mapping would also analyze who was using such commons-based strategies, and provide initial pointers for future political alliances on issues of patent and copyright policy. We now call this track the *Industrial Cooperation Project* (ICP). The second track would focus on developing new approaches to studying online cooperation on a much larger and more comprehensive scale than attempted in the past. We call this track the *Online Cooperation Research group.* (OCR)

We hired two fellows to direct the two parts of the project. Carolina Rossini is the project manager of the ICP. Aaron Shaw is the project manager for the OCR. Each has worked to develop a methodological framework, cooperative research infrastructure, and to hire and train researchers to assist with the research. The bulk of this report is made up of the progress reports from the two tracks, and includes a series of annexes that present parts of the methodologies developed and results achieved to date.

*The ICP*

The methodological approach of the ICP is well known and understood: the industry case study. The primary challenge here was to create a conceptual map that would allow us to: (a) standardize observations across sectors; (b) represent conceptually and visually the relevant attributes of players in each sector, and their changes over time, if any; and (c) identify whether industry practices have shifted toward more or less cooperative frameworks over time.

To achieve this Carolina developed a quadrant mapping approach. For the first year, we have focused on four major sectors that will likely have significant impact on global development and welfare over the long term: biotechnology; alternative energy; educational materials; and telecommunications.

The biotechnology sector has enormous potential implications for global health and food security; has well developed variability in practices, with some of the most proprietary alongside some of the most open and collaborative efforts. It therefore is a substantively important area and a potential model for our analysis more broadly. The second sector is one of enormous importance, but one for which there has been practically no work done on innovation policy—alternative energy. Here, the practices are less well developed, there is no real structure for commons-based practices, but there is substantial and interesting support from the current United States Department of Energy to embrace innovation-sharing practices as part of the global effort to address climate change and sustainability. In this regard, the area is important, and particularly fertile for developing new political alliances around questions of innovation and development as checks on IP. The third sector is central to education—that is, educational materials. This area is intermediate in the development of commons-based practices, counting large and historically powerful incumbents. The fourth sector, which we studied through an arrangement with an independent expert, was telecommunications, and in particular innovation in telecommunications. Here, much of the work that is open is done in standards setting, and the debates over telecommunications regulation take on some of the same characteristics as the debates over patents and copyrights in the other areas. Annexes II to V of the ICP progress report bring short but deep narratives detailing our findings for each sector, while annex I points out the questions that guided our course-plotting.

In the coming year we will deepen and conclude these initial studies, and extend the research to other sectors and actors in two main arenas of debate: food security and sustainability. Food security stands at the intersection of many disciplines, and actors addressing it face natural, market, and political challenges. The right to access food will face many barriers, but what is relevant for this study is the impact of practices around intellectual property in value chain of food production and access and in the diffusion of innovation in food. In this sense biotechnology innovations focused on food, GMOs, in parallel with farmers rights, traditional communities rights, among many others, are all components of this puzzle.

Sustainability is a second arena that deserves our attention. It is a hard concept to define, but in the environmental context, sustainability can be defined as the ability of an ecosystem to maintain ecological processes, functions, biodiversity and productivity into the future. For humans to live sustainably, the Earth's resources must be used at a rate at which they can be replenished. In the corporate world, sustainability is the idea of keeping production at levels that, while guaranteeing the long-life of a company, contributes for the conservation of the ecosystem that surrounds it. This concept includes "Green" technologies which cover a spectrum from alternative energy to resource consumption, increased recycling, decreased energy use or carbon output, and so on. Our

work here will start with a definition of segments we will focus on, and continue by deploying the research methodology we developed during the first year of the ICP project. This focus also motivates us to possibly work on a green-paper focused on climate-change, sustainability and energy, as a fuel to discussions under the new administration.

### The OCR

The Online Cooperation Research project is a multifaceted effort to analyze the broad space of online cooperation, collaboration, and commons-based production. As part of this project, we have begun work on several distinct avenues of research as well as the development of two web-based platforms for distributed and collaborative research.

Researching cooperation and commons-based production online poses multiple methodological challenges , and we present them in  greater detail in the OCR report. The major difficulty stems from our move away from the anecdotal and deep case studies that dominate the field, towards larger scale observational research from which we aim to draw more general conclusions.  Not only does no generally accepted definition of online cooperation exist; but also there is no census or comprehensive list of cases from which to draw a valid sample. To address these problems we have developed and tested completely new methodological approaches. As a starting point, we chose to use Wikipedia to derive a sample of cooperative systems online and to measure the validity of our sample in relation to other sites on the Net.   We also designed a method for observing a very large set of sites. To do this we needed either a very long time to observe, or a very large number of observers. We opted to develop a platform that would allow us to do the latter by creating an online tool that will harness a distributed workforce to answer questions about our selected sites. In addition, we wrote, revised, tested, and pilotedover seventy survey-style questions that we want to ask about each site. Finally, we worked with a team of researchers and collaborators to begin a series of smaller-scale comparative studies that examine cooperative systems in greater depth.

In conjunction with our research, we have also begun development work on a pair of open and collaborative research platforms online. The first of these is ScriptGen, a flexible, web-based research tool capable of distributing surveys, questionnaires, or experimental treatments and integrating seamlessly with the Amazon Mechanical Turk distributed labor market. The second platform is Coopedia, a collaborative repository of observations and analysis about cooperative and commons-based systems online. Both platforms will be licensed under the GPL and will be free for others to use, adapt, and re-purpose.

We expect the second year of the OCR project to involve three major elements. First, we will launch, collect data from, and analyze the results of the large-scale observation study—the largest such human observation of the Net attempted to date. Second, we will focus the members of our researcher team on in-depth analysis of selected sites, in order to complement the online observation arm of the study.  Third, we will develop, test, refine, and release versions of the two online platforms, ScriptGen and Coopedia.

## *Conclusions*

Overall, the year has been used to build a powerful research team, refine and apply new methods, and build new observational platforms. There is great enthusiasm and commitment to the project, and both arms of the study are proceeding on track to deliver what we believe will be genuinely new insights into the state of commons-based practices, both online and offline. These, in turn, will allow us to provide new intellectual heft and nuance to the debates over the relative efficacy and role of commons-based and peer production, and will improve our ability to map the political economy of the commons in the networked information economy.

# Contents

**Industrial Cooperation Project (ICP)**

# Contents

# Institutional Cooperation Project

**Progress Report for Ford Foundation**
**June/2009**

*I - Research Focus/Intro/Project overview*

The research project aims to map models of knowledge flow and appropriation across a range of economic variables. Our goal is to understand how intellectual property and other factors affect the movement and utility and usage of knowledge generally, through the selection of a targeted set of industrial sectors for research and analysis. We are mapping the actors, trends, and activities around cooperation in each of these sectors to understand the broader forces motivating cooperation more generally.

Our four sectors are biotechnology, educational materials, alternative energy, and telecommunications. We are interested in who the actors are in these sectors, what models they choose to leverage in their knowledge work, and why and when commons-based models emerge and gain traction. The project is built on an analysis of four disciplines representing a gradient of knowledge appropriation styles ranging from significant commons adoption to negligible commons adoption, and focuses within each discipline on paradigmatic knowledge products for deeper analytics.

These four sectors represent a strong cross-section of the contemporary economy. As such, they provide us a powerful lens to examine case studies and begin to develop theories about how, when, where, and why commons-based approaches develop, succeed, or fail in practice. We developed a mapping device inspired by the classic "Gartner quadrant" approach that allows us to map actors - including companies, cooperative projects, and so forth - into a standardized visualization of cooperation, as well as to map evolution of cooperation by actors.

The early results of the project are intriguing. We see from our examination of biotechnology the outlines of commons-based production in genomic data, as well as the impact of copyrights and patents slowing similar production in genomic tools and scientific literature. We see in educational materials enormous promise in collaborative knowledge generation for courseware and textbooks, but in alternative energy we find less evidence of spontaneous emergence of the commons. And in telecommunications we see the widespread emergence of commons-based systems ranging from contracts to technical protocols.

The report ahead lays out methodologies and introduces our version of the quadrant, and contains detailed annexes providing deep information on our four sectors. Three annexes – biotechnology, educational materials, alternative energy – were written as part of the

core research project, while a fourth – telecommunications – was written by an expert in the field. Each annex can be considered an introductory mapping of the field in our context. The project also lives in the research wiki, which contains a wealth of accumulated information supporting each annex. Finally, this intermediate progress report sketches long-terms goals for future directions that the project might take.


## II – Research Team:

**Host Institution:** Berkman Center for Internet and Society at Harvard University

**Lead Professor:** Yochai Benkler

**Research Fellow:** Carolina Rossini

**Research Assistants:**
Alternative Energy: Silas Bauer; Andrew Clearwater;
Biotechnology: Brendan Ballou;
Educational Materials: Erhardt Graeff

**Telecommunications Expert:** Michael Steffen.

*Team curriculum information available at:*
https://cyber.law.harvard.edu/commonsbasedresearch/ICP_Team


## III – Research Justification and Methodology:

The initial research questions that motivated this study are: How are components of the industrial structure changing in how they deal with and manage knowledge embedded assets, in different industries, different business models, and different sets of actors? How (and if) are they incorporating commons-based strategy? Thus, the Institutional Cooperation Project focuses on understanding how institutions shape the kind of organization available for sustainable human cooperation (social, economic and political behavior).

This interest is also justified in order to understand in each extent and how the Intellectual Property System is a need or is operated in different sectors. As Benkler wrote: "If some information producers do not need to capture the economic benefits of their particular information outputs, or if some businesses can capture the economic value of their information production by means other than exclusive control over their products, then the justification for regulating access by granting copyrights or patents is weakened" [1]

---

[1] Wealth of Networks, p37

The project was structured in three main phases regarding (1) the understanding of a sector configuration and mapping of actors and market practices regarding Intellectual Property in a chosen economic sector (Field); (2) the identification and analysis of paradigmatic commons-based cases; and (3) the characterizations of common practices for taxonomy development.

## General Methodology Framework
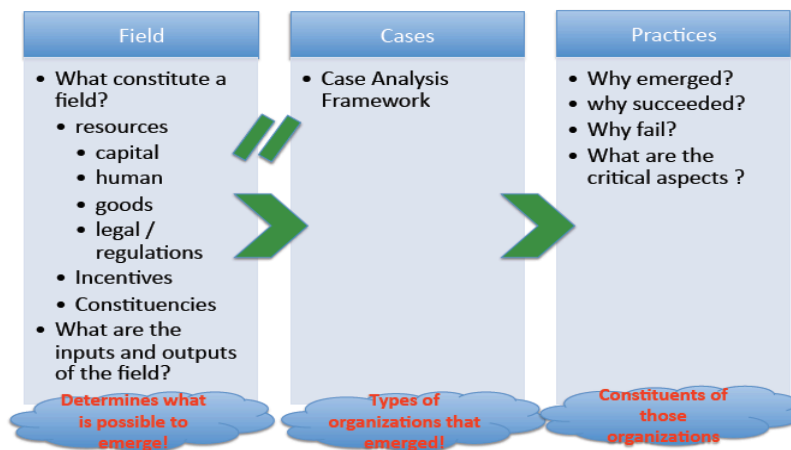


The phase (1) was carried on mainly through literature and market reports review, and interviews with sector experts, scholars, activists and investors. The research framework that guided this first phase research is attached as **Annex 1** to this report.

*IV – Mapping Device*

The word "commons" refers to a particular form of structuring rights to access, use and control resources. The mapping device we developed is focused on understanding how actors in a chosen sector structure exactly the access, use and control of resources in terms of participation and regulation as below defined.

The intent is to categorize the "commons" in a more rigorous fashion, so that we might understand its different implementations in different sectors, its constituent parts, and perhaps even arrive at an "atlas" or "taxonomy" of the commons that we can use in future design and evaluation.

Most of the cases we will examine are knowledge resources (ranging from explicit knowledge like textbooks to tacit knowledge encoded in biological tools). As such, understanding their usage systems requires an understanding of the forces that governed their creation, since this frequently influences how the structure of access, use and control of resources get settled.
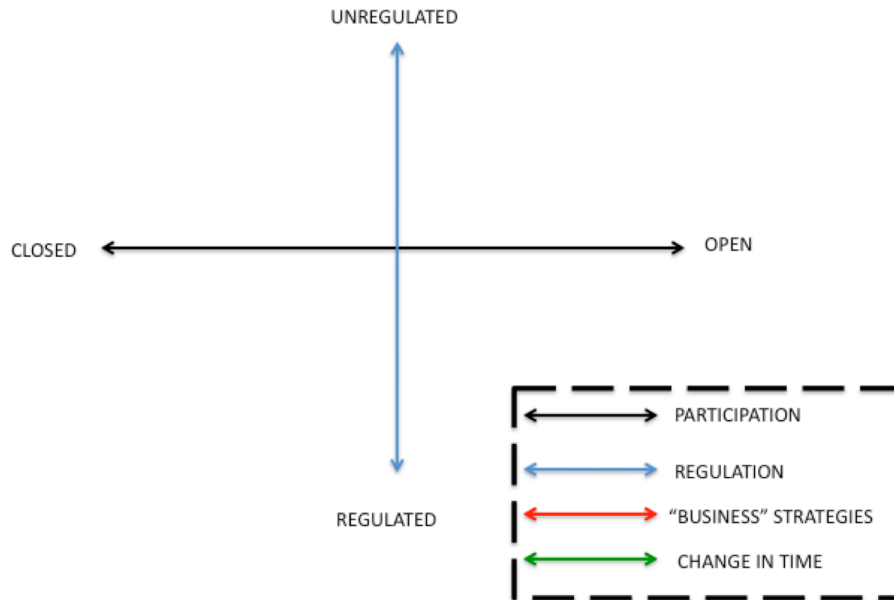
Thus, through placing actors in the quadrant and capturing their moves over time, we learn how the intellectual property systems work de facto in relation to a specific knowledge resource.

Specifically, regarding the mapping device, the axis of Participation deals with the question of "how open it is to join" either as a creator or a user. From Closed to Limited to Open, participation is the measure of the constituency – it deals with the person herself and not the issue of the person's freedoms.

The axis of Regulation deals with the quantity and quality of norms that frame the rights of a user in any certain Case. In this sense, any thing, from norms, to contract, to law, to the practice of non-enforcement is measured as a set of elements that constitute and determine how a certain environment will behave. It may vary from Regulated to Limited to Unregulated levels.

The questions are focused on finding metrics through objective answers that are later rearranged in groups that denote characteristics such as symmetry, freedom, predictability, openness, transparency, collaboration, user-integration and autonomy.

*V – Research Platform:*

Taking into consideration the geographic distribution of the research team and also the long-term goals, such as providing a platform for future learning opportunities, we adopted a wiki platform as the base for our research and also as a way to document the research progress.

This platform is available at:
http://cyber.law.harvard.edu/commonsbasedresearch/Main_Page

We believe that this experience can also provide us inputs on how best to organize a platform for collaborative knowledge production in the long-term run of the project, in addition to helping us organize our research progress.


**VI - Main Sector Progress:**

Project narratives for each sector are attached as annexes. Raw research material, links, case studies, and other information including detailed bibliographies can be found at the wiki pages linked below.

1) Alternative Energy
https://cyber.law.harvard.edu/commonsbasedresearch/Report_May_2009#Alternative_Energy

2) Educational Materials
https://cyber.law.harvard.edu/commonsbasedresearch/Report_May_2009#Educational_Materials

3) Biotechnology
https://cyber.law.harvard.edu/commonsbasedresearch/Report_May_2009#Biotechnology.2C_Genomics.2C_and_Proteomics

4) Telecommunications
https://cyber.law.harvard.edu/commonsbasedresearch/Report_May_2009#Telecommunication

*VII – Research Timeline:*

This timeline focuses on the main goals of each research semester for the Institutional Cooperation Project. Specifics regarding each of economic sectors under analysis are provided in the annexes to this report.

*First 6 months of the project:*

This sub-division of the Cooperation Project started in late 2008. The first couple of months of the project were focused on discussions on the research methodology and choice of initial sectors. Soon after, the research on the chosen sectors began. The non-structured interview process started in late March. Finally, Carolina Rossini presented two research talks during discussion groups: one at the Berkman Cooperation Seminar (March) and a second one at Yale Law school (April).

***Future work:***

*Middle Term: June-December 2009*

The future work can follow two possible strategies currently under discussion:

• Increase vertical sector research, developing, for instance, case studies of commons-based cases and developing further mapping analysis. In this regard, specific suggestions by area are provided in the respective annexes.
• A second and complementary activity to expand the research horizontally, including new economic sectors in our research.

*Long Term: January 2010-June 2010.*

• Cross-sector comparison of the similarities and differences of the conditions that allow for the emergence of commons-based models in the economic sectors under study.
• Development of commons-based cases taxonomy.
• Final organization of the wiki platform for publishing and opening to community contributions and use.

***VIII - Annexes:***

Annex 1 – Research Skeleton
Annex 2 – Summary findings and next steps in Biotechnology
Annex 3 – Summary findings and next steps in Alternative Energy
Annex 4 – Summary findings and next steps in Educational Materials
Annex 5 – Summary findings and next steps in Telecommunications
Annex 6 – Sage Paper

**ICP PROGRESS REPORT ANNEX 1**

**Research Skeleton**

Part 1. An introduction providing an overview of economics of Intellectual Property in the sector under analysis, with the objectives of:

 1.1. provide a literature review of the IP debate in the sector. (Where and how IP is working or not, and what are the relevant topics around IP in this sector?)

 1.2. indentify what are the other innovation incentives mentioned.

 1.3. find data on "how much of an increase of the tendency towards enclosure" characterizes the sector. (e.g. How much has patenting increased over time? Which actors cab be identified pushing this trend?)

Part 2. Provide an overall picture of the sector:

 2.1.  How was this field born and how is it evolving?

 2.2. What are the main business models?

 2.3. What are the innovation dynamics in this field? (inputs/outputs, cycles of innovation/ disruptive or incremental innovation?)

 2.4. How does knowledge flow in this field?

 2.5. Is this field replicating models from other fields?

 2.6. How many companies and how concentrated is the sector?

 2.7. How much money do they make or how much money do they "move" in the American economy?

 2.8. How important is research from universities in this specific field?

 2.9. How important is public funding in this field?

 2.10. How important is private funding / venture capital in this field?

 2.11. Are there any specific public policies (from agencies, federal or state policies) that give incentives for openness or enclosure?

 2.12. What is the cost structure of the field?

 2.13. Who are the producers, the buyers, and the users?

 2.14. What is the structure of power from the production side and what is the structure of power in the demand side? (E.g., who has the power to control production and demand? How is the control distributed? How is the relation among producers, adopters, buyers, and users? Do these relations bring any market dysfunctions?)

Part 3. Define what kinds of market-segments are relevant and how the knowledge inputs for innovation in a certain sector are characterized. (e.g. In Biotechnology, we adopted the division of data, texts and tools as the main knowledge inputs and outputs in different moments of the biotechnology value chain)

Part 4. Define the main legal tools of protection (privatization) available for the field (patents, copyright, trademark, trade secrets, contracts, public domain) and the trends regarding their enforcement. Develop a matrix exercise on the trends and how they impact in openness or closedness of strategies.

Part 5. Define the competitive advantages in the field and the barriers of entry, and how this may impact in the emergence of common-based models.

Part 6. Identify the biggest for-profit companies in the sector (the sample should be of 10 firms maximum, based on their market share, their incumbent position and their importance in shaping the sector)
       6.1. How is the market distributed?
       6.2. Where are they located? Are there any incentives for specific locations?
       6.3. Correlate them with their main outputs and market segments.
       6.4. Correlate them with IP strategies.
       6.5. Indentify how (and if) they contribute to the commons and if this is a "experimentation" or a clear "adopting" commons-based approaches
       6.6. Identify these cases and treat them as entities that will also be placed in our mapping device (the quadrants)
       6.7. Try to understand which companies use IP to enclose knowledge or to open knowledge as parts of its innovations strategies

Part 7. Identify the biggest non-for-profit institutions (public or private) in the sector (Repeat analysis of item 6 for item 7, adapting when it is needed)

Part 8. Indentify the (5) top Universities in this field in terms of importance of R&D and innovation focused on the chosen sector.
(Repeat analysis of item 6 for item 8, adapting when it is needed)

Part 9. Identify industry/universities/professional associations and civil society organizations that shape the IP discourse in the sector. Identify their policies, lobby, recommendations and/or best practices related to IP of their members.

Part 10. Identify commons-based cases or alternative open business models that emerged in the sector
(Repeat analysis of item 6 for item 10, adapting when it is needed)

Part 11. Identify industry/universities/professional associations and civil society organizations that shape the IP discourse in the sector. Identify their policies, lobby, recommendations and/or best practices related to IP of their members.

Part 12. Indentify and analyze private foundations that give grants in this field and how they shape the IP discourse.

**ICP PROGRESS REPORT ANNEX 2**

**The Biotechnology Sector**

## I - Introduction

Biotechnology is a broad term that refers to technologies that relate to the modification or understanding of living organisms and biological systems. In its broadest sense, biotechnology can include traditional practices such as cross-breeding and organism-based products like cheese. However, its common usage is much more specific to the research and businesses that leverage modern techniques to identify genetic material, human genetic variation, and the relationship of genetics to diseases and potential therapeutic interventions. The products of biotechnology include scientific papers, foundational data like gene sequences, data "sets" representing the output of experiments, biological tools, computer software, diagnostic kits, therapeutic compounds, and much more.[2]

There is a broad set of actors in biotechnology ("biotech"). Funding comes from massive investment by governments as well as from private research foundations, venture capitalists, and public markets. Research takes place in academic settings, industrial research laboratories, and government facilities. Businesses leverage biotech from the earliest stages of gene investigation all the way through to pharmaceutical product development and investigation.

Beyond these traditional actors, the material of biotech itself covers a wide range. Molecules under investigation range from simple nucleic acids to complex, 3-D protein structures. Organisms of study scale from simple (yeast, flatworms) to the most complex (the human body). And at each stage of biotech, a dizzying array of tools can be developed, deployed, researched, monetized, protected - or be tools for cooperation.

## II - Research Focus

Biotech offers a rich history of cooperative innovation and knowledge development. The field is marked in many areas by widespread cooperative behavior. Within biotech, the fields of genomics[3] and proteomics represent an ideal focus for an analysis of cooperative behavior and commons-based knowledge generation - there are long-established actors, projects, and cooperative systems, covering most of the classes of products produced by

---

[2] Biotech also spurs innovations in areas such as food, energy and in other sectors. However, for the purposes of this phase of this research, we are focus on the biotech that provides inputs to the pharmaceutical sector, which provides a rich field of coherent cases to study.

[3] The science of genomics is focused on 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 to determine the activity and function of genes. Genomic scientists produce genetic, pathway, and functional information analysis about the genome in attempts to place the DNA code in context inside living beings, typically in order to improve human health or advance agricultural technology.

biotech, and across a wide range of tools and knowledge. Good examples of these projects and systems include the Human Genome Project, the HapMap, and distributed genome annotation. There are also well-formed markets including both academic and corporate players in bioinformatics, scientific publishing, and research tools dissemination.
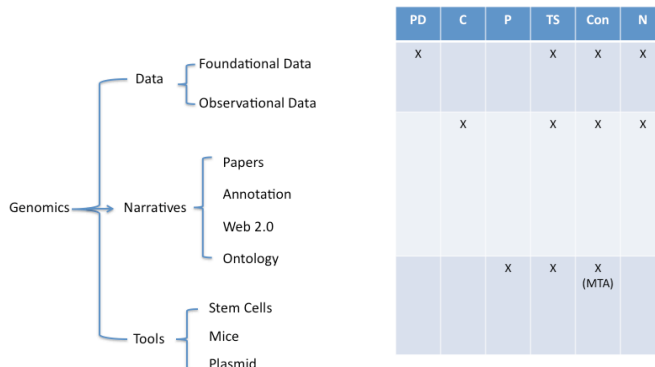
The entire fields of genomics and proteomics have been hit by a series of waves of disruptive technology, with new waves often hitting before the last wave is fully integrated. This creates a pressure to adopt technologies quickly and adjust behavior (both cooperative and competitive) in a "real time" fashion, which in turn provide this research with cases where we may not only identify a cooperative arrangement, but watch the evolution and development of cooperation over time. This unique aspect of genomics and proteomics is captured and mapped in our Quadrants tool, in which we examine the development of cooperative arrangements in three biotech output areas: narrative text (primarily scientific publishing), foundational data (gene and protein sequences), and research tools (biological materials, from simple to complex).

### III – The Intellectual Property Factor

The wide range of biotech outputs means that the entire range of intellectual property rights come into play at some point in the cycle.

Copyrights, which govern the right of users to make and distribute copiesof creative works, attach to journal articles, conference proceedings, posters, blogs, wikis, and other narrative forms of communication. 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 owner to extract revenues from subscribers.

Copyrights also factor into 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" 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.

| | PD | C | P | TS | Con | N |
|---|---|---|---|---|---|---|
| Data — Foundational Data / Observational Data | X | | | X | X | X |
| | | X | | X | X | X |
| Genomics → Narratives — Papers / Annotation / Web 2.0 / Ontology | | | X | X | X (MTA) | |
| Tools — Stem Cells / Mice / Plasmid | | | | | | |

*PD=Public Domains; C = Copyright; P = Patents; TS = Trade Secret; Con = Contracts, including Material Transfer Agreements; N = Community construed norms.*

Patents, which govern the rights to practice an invention, attach to innovations like genetically modified sequences, disease mechanisms of action, engineered tools like stem cells, mice, 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.


### IV – Segment Focus: Genomics and Proteomics Publishing, Data and Tools

Within genomics and proteomics, 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, to the impact of a natural public domain data status, to the impact of community norms, legislative and judicial changes on research tools, our study of tools, narratives, and data provides a lens onto cooperation across most of the key actors and segments of the industry.


### V - History of Biotechnology in the USA

The history of the biotech business can be traced back to 1953 when Crick and Watson discovered DNA. But commercialization of genomic discoveries did not begin until 1973 when Stanley Cohen and Herbert Boyer invented the rDNA process, by which foreign genes could be inserted between the ends of existing genes. Boyer and venture capitalist Robert Swansen used their patented rDNA process to found Genentech, the first biotech company, in 1976. The 1970's and 80's saw a proliferation of biotech companies: some, like Genzyme, Amgen, and Genentech were successful; others were not. Because none of these firms had marketable products at the time, they grew through outside investment - and also dependent on intellectual property (the only metric against which to measure progress and value at that stage).

This outside investment grew and shrank in cycles, forcing many unprofitable firms into bankruptcy, and others into deals with larger multinational companies. Today there are over 330 publicly-traded biotech companies (Ernst & Young "Beyond Borders" pp. 19), of which the eight largest bring in over USD$35 billion. And though the size of the industry may have grown, the overall dynamic of the market remains much the same as it

was thirty years ago: only a handful of the largest companies ever bring products to market, and most biotech firms survive through either public or private equity sales.

Within biotech, genomics and proteomics experienced a massive boom in the 1990s, driven by the massive government investment (US$3,000,000,000) in the Human Genome Project, a 13-year effort to sequence the human genome led by the U.S. Department of Energy and the U.S. National Institutes of Health. Complementary efforts to sequence less complex genomes such as the fruit fly and the mouse were also part of the HGP. Most of the work was done in academic locations in the United States, Canada, New Zealand and Britain, with scientists from China, France, Germany, and Japan also key members in the consortium. Celera, a private company, raced to create a private version of the genome, driving the government funded effort to work faster, and also pressuring the scientists involved to develop more robust strategies for collaboration and cooperation (known as the "Bermuda Principles"[4]).

Out of the HGP came thousands of papers and patents, and an explosion of startup companies in genomics and proteomics, which peaked in the genomics "bubble" of the late 1990s. There was great expectation that genomics companies would dominate the new face of drug discovery and development, which faded as the bubble burst and hundreds of companies went out of business. The publication of the complete human genome in the public domain also had a significant impact on the potential for companies to use trade secrets to protect their data products.

At the same time, major new "big science" projects like the HapMap database of human genetic variation were begun, and the remnants of the bubble companies were absorbed into larger companies or turned into more open-source approaches. Taken together, this part of the biotech history has dramatically affected cooperation in the data sector, as we'll examine closely in the quadrants.

The business of biotech was also dramatically affected by changes in the IP environment. Court cases allowing the patenting of modified genes, and legislation encouraging universities to acquire and license patents on government-funded research, paralleled the genomics bubble with a dramatic increase in the number of patents filed by universities on biology. The importance of biological materials grew 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. We will examine this in the tools quadrant mapping exercise.

---

[4] The Rules were simple and clear: all data was in the public domain, and it would be posted online with 24 hours of coming off the machines. However, scientists using the data were expected to check and see if the data had been "published" yet (the fuzzy part) and if it was unpublished they were expected to honor some norms about the data. See here http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml and http://www.sciencemag.org/cgi/content/summary/291/5507/1192

| Number of Biotech patent approvals (Oliver pp. 59) | | | | |
|---|---|---|---|---|
| | Drugs | Microbiology | Multicellular organism | Recombinant DNA |
| 1977 | 660 | 591 | 0 | 14 |
| 1982 | 730 | 711 | 1 | 111 |
| 1987 | 958 | 1099 | 19 | 204 |
| 1992 | 1691 | 1965 | 52 | 356 |
| 1997 | 3372 | 4178 | 318 | 506 |

At the same time, the scientific scholarly publishers saw an increase in their ability to leverage technology. Similar to the music and movie industries, publishers 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, 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. In reaction to this pricing crisis, the Open Access movement evolved from a relatively discipline-specific affair into a fully empowered active system in the narrative space. The quadrant examining narratives will address these trends.

## VI – About the Biotechnology Market

After the bubble, genomics companies trended towards massive-scale sequencing, cloning expertise, and functional genomics. Current business models in biotech bear little resemblance to the bubble models, which focused on building proprietary databases of genetic information and large-scale patent filing projects. Most companies that have survived have done so through a radical shift towards the discovery, trial, and marketing of pharmaceutical products - the platforms that once formed the core of their business now represent just a portion of the overall drug discovery process. By 2006, there were 336 publicly traded companies in the US in 2006, but the eight largest Biotech companies bring in $35,821,000,000 out of a total market revenue of $55,458,000,000, accounting for 65% of the industry's revenue.

The pure genomics and proteomics industry today is instead dominated by genomics-based drug development companies, toolmakers, bioinformatics software providers and information companies. The shift from the old models to the new models has been paralleled by the emergence of open source databases of genomic and proteomic information in the public domain.

The market for research tools is harder 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. The market for tools is also affected by the establishment of biological resource centers (BRC) into which funders sometimes mandate deposit of research tools developed under funded research.

The market for narrative text in genomics and proteomics exists in an almost completely unrelated plane, and is dominated by a small number of publishing companies like Elsevier, Nature, Springer, Wiley-Blackwell, and others. It is marked by steadily increasing prices and strategies in which less-valuable journals are bundled with higher value journals, and by a powerful reliance on copyrights. Recent changes in the market have been driven by the emergence of Open Access publishing and "author self-archiving" in which the cost of article publishing is paid at the end of a research cycle, instead of the beginning. A 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.

### VI.1 – Production and Distribution Cycle

Production and distribution cycles in genomics and proteomics differ wildly depending on what is being generated: data, narrative, or tools. 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. 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 scientist, and popularity of the material.

Other tools not studied in this report under the tool category (because they are not biological materials) 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 tends to depend on technical capacity and class of data - if the Entrez system accepts deposit of a class of data, distribution of that data tends to be dominated by Entrez. Otherwise, there is a wide range of distribution behavior ranging from posting of data files on laboratory websites to "email me to ask for the data" to outright withholding.

Production and distribution of narratives exists at a different cycle, 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. 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 pre-print articles, and hyperlink articles into websites, emails and more. The ability of individuals to circulate their work marks a major change in narrative distribution.

## VII – Policy Trends

Policy trends in genomic and proteomic research unsurprisingly trend towards 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 across Europe, in Australia, and elsewhere.

A similar set of trends are 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. At the smaller scale, the experimental data generated in laboratories is subject to less formal policy requirements though the expectations, informal and formal, are tilted towards sharing, with NIH grant proposals requiring the submission of data sharing plans.

Tools 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 Bayh-Dole as an influence for universities to patent research tools. Bayh-Dole is also occasionally brought into play in the sharing of data from which patentable inventions might be drawn.

## VIII – Funding

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. 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.

*IX - Secrecy vs. Openness >>> Cooperation Mapping Quadrants*
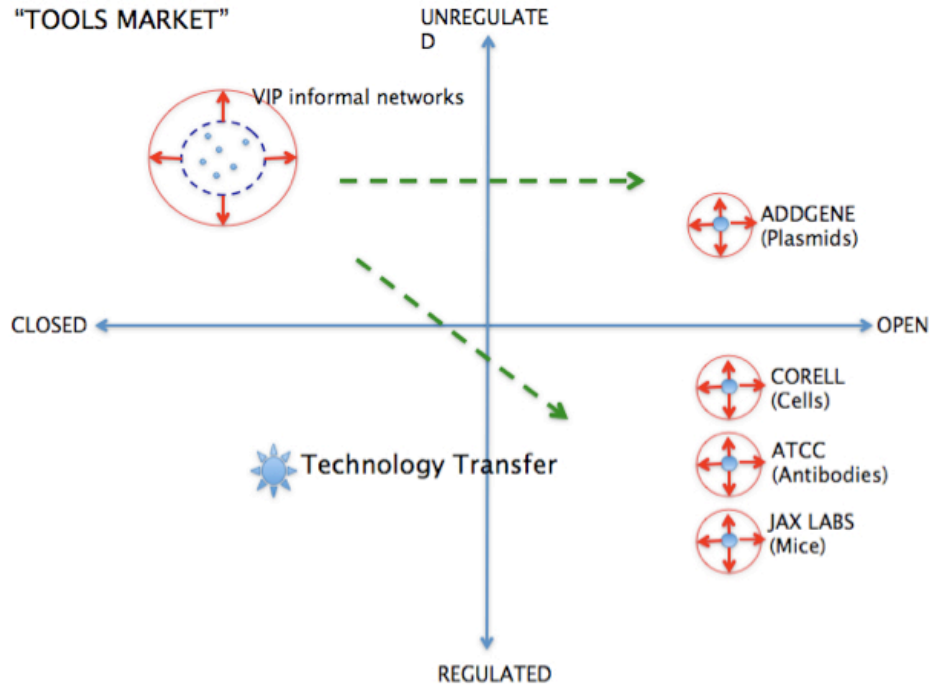


In our quadrant mappings, the tools market has a major concentration in the unregulated-closed sector, representing the traditional methods by which tools move around between researchers, which we call "VIP" (an aphorism for "vial in pocket"). Under the VIP system, researchers contact other researchers by phone (now email) and simply request the material. The scientist providing the material must assign resources to manufacture a copy of the material, and the technology transfer offices of both institutions must come to legal agreement on transfer terms. This is in addition to the market pressure of funding competition that creates instincts towards competitive withholding; creating a system that is not regulated in the traditional sense yet trends towards closure. The green lines indicate the influence of BRCs on tools, as the existence of standard intermediaries and semi-standard contracts, with disclosed catalogues on the web, increases the openness of the system.

Addgene is a BRC for plasmids, which typically are felt too unimportant to merit the attention of technology transfer, placing it in the upper right quadrant, while three other BRCs sit in the lower right, indicating the importance of patents on the various classes of

tools provided, and the degree of regulation increases as the patents gain importance. One BRC, Jackson Laboratories manages genetic modified mice on behalf of the US government and other research institutions. The case of mice and how the change in norms of accessibility impacted in innovation is documented in the research Wiki Platform[5]. The blue star in the regulated-closed sector represents the Wisconsin Alumni Research Foundation, which manages some of the most important patents around stem cell production and aggressively protects the technology, but has been moved by the external markets to offer more open and less regulated research access to the technology, as a paradigmatic case illustrating the complexity of mapping an individual tool.
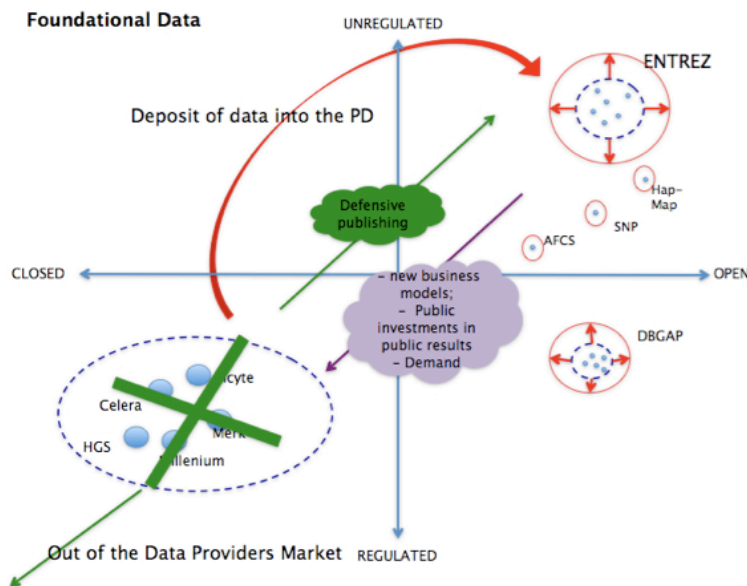


The foundational data quadrant mapping illustrates many of the elements mentioned in the markets discussion. Entrez, in the upper right quadrant (unregulated-open) is dominated by the Entrez resource. The Entrez Global Query Cross-Database Search System allows users to search across dozens of life sciences databases housed at the National Center for Biotechnology Information inside the US National

Library of Medicine, all of which are legally in the public domain. Entrez includes the full genome sequences of the Human Genome Project and many other resources. The legal status and technical searchability provided by Entrez is a powerful force towards unregulated openness - no registration is required, no data collected, no licenses signed.

Over the past ten years, companies like Celera, Incyte, Human Genome Sciences,

---

[5] See here:
https://cyber.law.harvard.edu/commonsbasedresearch/Give_an_overall_picture_of_the_BGP_field#Tools
and here
https://cyber.law.harvard.edu/commonsbasedresearch/Data%2C_narratives_and_tools_produced_by_the_BGP_field#Mce

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 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 unregulated-open space and creating market standardization around the Entrez technology platform as well as the public domain.

Other projects to map human genomic variation (HapMap, SNP consortium) and cellular signaling (Alliance for Cellular Signaling) align with the Entrez model and frequently are absorbed into it. On the regulated-open quadrant we place the dbGAP project, also part of NCBI, which contains data that correlates genes to function, is legally in the public domain and available to qualified researchers with a user account at the NIH, but requires the submission of a data access plan to a review committee, which introduces a degree of regulation otherwise not seen at NCBI. This process was spurred by the realization that patient data inside dbGAP could be identified using complex algorithms, which led to the control over data via secrecy protections and access controls.
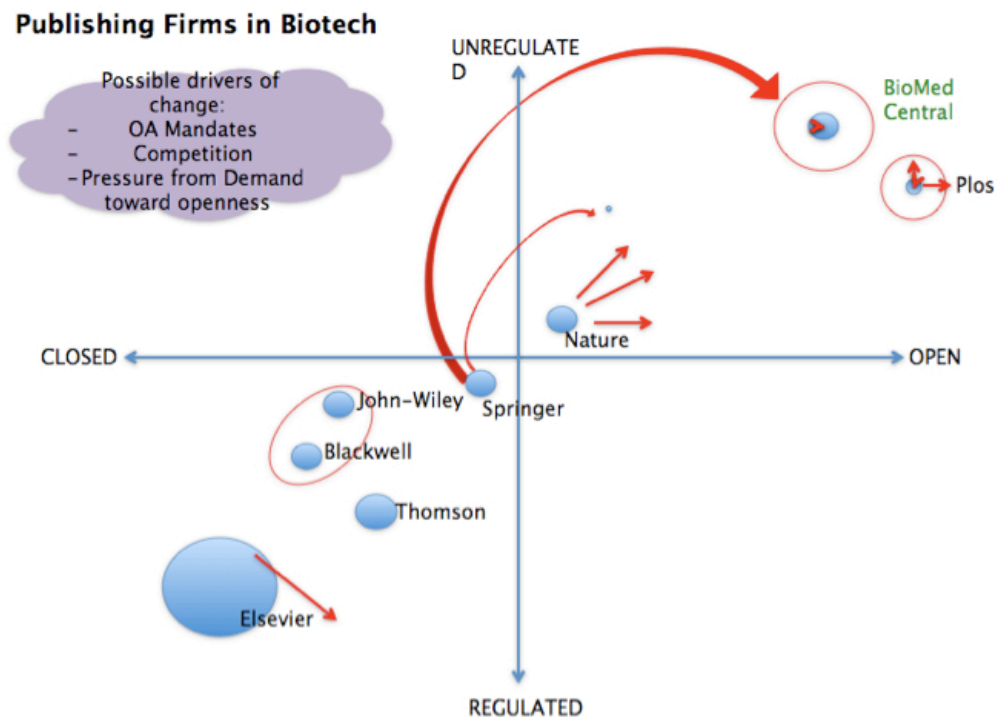


The markets for publishing narrative text in genomics and proteomics exists historically in the regulated-closed quadrant, with subscription-based print journals sold primarily to libraries and funded scientists the dominant paradigm until the advent of widespread

Internet in the late 1990s. To the extent there was openness it was enabled by the choice of the subscriber to allow browsing of physical periodicals in open libraries at universities or public libraries. Copying was done on photocopy machines and lightly controlled. With a move to electronic access to journals, publishers moved towards license-based "leasing" of the scholarly literature rather than sales, eliminating significant rights previously held via subscriptions purchase (such as rights under the first sale doctrine, and fair uses blocked by technical protection measures and rigid standard contracts).

Elsevier is the dominant publisher in the field, with approximately half the market in genomics and proteomics journals, and a set of other publishers including Wiley, Blackwell, Springer, Nature, and Thomson form the rest of the traditional market. In the quadrant, we place Thomson as a larger entity due to its control of the ISI "Impact Factor" rating system for journal quality, which gives it disproportionate market power, and slightly more open than other journals as the impact factor itself relies on the existence of at least somewhat open systems, though it is itself highly regulated. Springer has a smaller presence in biotech than in other scholarly disciplines but was an early publisher to offer to authors the right to retain significant rights to their work through the purchase of the right to distribute works under Creative Commons BY-NC licenses (at a cost of US $3000 per article). Wiley and Blackwell joined forces via merger, bringing a plurality of scholarly society publishers into a larger publishing entity; the societies do not as yet speak with a common voice on issues of openness and regulation and the combined entity is so far relatively cautious, meriting its placement as well in the regulated-closed quadrant.

Nature is the only traditional publisher that we place in the unregulated-open quadrant, though quite close to the central connection point. Nature has experimented significantly with Web 2.0 technology via the Nature Network, the Connotea social bookmarks for life sciences, and a variety of discipline and disease-specific "gateways" integrating information. Nature also publishes a small number of new, online-only journals under the Creative Commons BY-NC license and committed in 2008 to publish all papers describing the initial complete genomic sequence of an organism under BY-NC as well.

The open-unregulated quadrant is dominated by the Public Library of Science and BioMed Central (PLoS, BMC). Although BMC has almost 200 more journal titles than PLoS, the latter has an outsized impact in the market due to its high profile leadership including former director of the NIH and Nobel Laureate Harold Varmus. PLoS is a non profit organization and is moving to develop a novel method of publishing called PLoS One, in which articles are rapidly reviewed for scientific veracity only, and "impact" emerges over time from citation indices, comment threads, and other methods. PLoS One is already disrupting the market by enabling PLoS to approach financial sustainability. BMC's revenues approached US $15M in 2008 and established BMC as the clear leader in for-profit companies in the unregulated-open space, leading to its acquisition by Springer in late 2008 (visualized in the quadrant by the red arrow linking the two companies).

***Defining the "clouds"***

We use clouds to represent less-defined forces on openness and regulation. In the data quadrant mapping, we use the cloud to show the impact of corporate publication of data as a defense against patents (preventing tie-up of downstream drugs from patents on early-stage genomics, an ironic echo of what one finds in many public discussions about public health and the role of universities to leverage their "IP to open up drugs"), as well as to note that new business models in data services, data "cleaning," bioinformatics based on open unregulated data, and the concept that public investment should create public goods. In the publication quadrant we represent the impact of OA mandates in the cloud pulling towards open unregulated as well as the impact of competition from open publishers on closed (a two-way pressure!) and the pressures imposed by a growing desire by taxpayers that their investment in research lead to accessible articles.

## *X - Trends in the field*

Foundational data appears to be settling in the open-unregulated quadrant. The combination of the native public domain status of raw factual data in the United States, the US investment in creating raw foundational data, the Entrez infrastructure, and the desire by large pharmaceutical entities to create precompetitive spaces in genomics and proteomics represent powerful market forces towards open-unregulated. Publishing narrative text also is subject to strong pressure towards unregulated-open by the US NIH Public Access Policy, though there is a secondary strong pressure being exerted by Elsevier and other publishers aiming to create open-regulated systems inspired by Facebook and Apple's iPhone store, in which public content from the unregulated-open space is integrated with proprietary content and a developer community builds applications that only operate on the combination content, which is only available to subscribers.

Tools are harder to characterize - less complex tools like plasmids are clearly moving towards the open-unregulated, and funders increasingly are focusing on tool access as a condition of funding, but the influence of patents and university technology transfer systems represents a powerful source of movement towards regulated-closed, or regulated-open instead. Movement towards "one-click" access to tools via BRCs and new methods to track and reward the deposit of tools may create an influence towards unregulated and open, but the market is still evolving.

## *XI - Next Steps*

In the next steps of our research we plan to:

- Look more closely at the distribution of regulation and openness across different classes of data and tools.
- Put more emphasis on case studies of certain actors in the genomics and

proteomics sectors that may exemplify the open or closed models that the research is based on.

- Possibly add surveys for specific topics of importance like Biological Resource Centers and standardized data formats.
- Follow the activities and conduct a case study of Merck's Sage initiative and other pharmaceutical moves to share data and tools as examples of collaboration on genomics and proteomics.
- Conduct more research on the US National Institutes of Health to understand how they develop new policies and procedures like the Open Access policy and the reports on research tools, how they develop standard technologies like Entre, and how their practices impact in the emergence of closedness or openness arrangements.
- Explore the other sectors of biotechnology, with a move towards the use of biotech in the pharmaceuticals, energy, and food industries.

**ICP PROGRESS REPORT ANEXX 3**

**The Alternative Energy Sector**


*I - Introduction*

The market for alternative energy technologies in the United States has grown due to a myriad of indirect and direct factors. Indirectly, global climate change concerns and volatile fossil fuel prices, along with US energy security concerns tied to its dependence on unstable foreign sources of oil, have pushed alternative energy into a strategic position of importance. Direct factors affecting growth have been a recent increase in private funding for alternative energies, and a growing public-sector opinion that supporting these technologies is in the best interest of the country. Up to this point, the US has lagged behind other countries, mainly those in Europe, in terms of both its public R&D and demonstration funding (supply-push policies), and its technology deployment funding (demand-pull policies).[6]

Though the US is the world leader in installed wind energy capacity and within the top 5 among most other alternative energy technologies, the amount of federal funding it has devoted to alternative energy R&D has declined by half since its peak in 1980. Private R&D funding has had a similar decline. Demand-pull policies in the US have been present since the late 1970's, but have suffered from ineffective administration. Most of these policies have been introduced on the state level and can vary widely throughout the country in their funding levels. Federal tax credits for alternative energy projects have been present since the late 1990s and can be a critical aspect in successfully funding new projects. The tax credits have also suffered from short periods of administration – two to three years – and occasional expirations.  The results have been boom and bust periods of alternative energy development in the US that coincide with the effective years and expired periods of the tax credits. While alternative energies provide a public benefit by reducing the environmental public health and climate impacts of carbon emissions, and providing new jobs, their high price point deters consumers, even those who may be motivated by altruism. It is evident that these supply-push and demand-pull policies are necessary for the growth of the alternative energy market due to the high cost of these technologies in comparison to conventional fossil fuel based energy sources.

---

[6] Gallagher, K.S., Holdren, J.P. & Sagar, -.A.D., 2006. Energy-Technology Innovation. Annual Review of Environment and Resources, 31. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1081940 .

## II - Research Focus

The Alternative Energy (AE) field represents a unique case for studying the trends in an emerging market. Some of the technology can be considered mature, however many are the barriers – technical, political or related to funding – that justify a young market.

Spurred in part by the increasing momentum of the "Clean Tech" movement, AE producers, consumers, and various regulatory and advocatory bodies are each responding to and evolving with the field, and thereby creating new market demands and offerings. While these trends are complicated in their economics, politics, and other social factors/barriers, the gradual consolidation of the field's largest producers is already perceptible in the wind market, for instance.

## III – The Intellectual Property Factor

Patents represent the most significant intellectual property tool involved in this field. Until recently, the IP factor did not parallel the usual IP debate found elsewhere regarding to access, sharing or balance. Many IP issues did not came to the center of attention of IP observers or even civil society groups focused on IP issues and development, because the debate over clean and renewable tech has been politicized and linked to long-term discussions around climate change, but not linked innovation and IP as in other fields like pharmaceuticals, software, and cultural works. In this sense, political strategies from Clean Tech and Renewable Energy industry associations were much more focused on policies to foster the adoption of these technologies over oil-based energy as explained in this report. Thus, we observed a stage in our research where few in the international IP community paid attention to the crescendo of patents in the Renewable Energy market.

However, this situation changed dramatically in the spring and summer of 2009 with the advent of the Obama administration making public statements about sharing technology related to energy. In reaction, the United States Chamber of Commerce, a leading lobby representing businesses, is expressing growing concern that moves to spread new energy technologies to developing countries could erode the intellectual property rights that have driven commercial efforts to innovate for generations. Late in May 2009, that group and representatives of General Electric, Microsoft and Sunrise Solar gathered in Washington to launch the Innovation, Development & Employment Alliance, or I.D.E.A. The initiative is aimed at pressing Congress and the Obama administration to ensure that global climate-treaty talks don't weaken protections on who can profit from new technologies that provide abundant energy without abundant pollution.[7] The creation of I.D.E.A. has been widely noted, with some alarm, in the IP "watchers" community, and likely means the status of AE as a less-observed IP sector is finished for good.

---

[3] http://greeninc.blogs.nytimes.com/2009/05/21/clean-energy-and-intellectual-property/
and http://greenpatentblog.com/2009/05/17/new-alliances-big-idea-strong-ip-is-essential-for-green-innovation/

*IV – Areas of focus: Wind, Solar and Ocean technologies for energy generation*

Our research has focused on the innovations in the alternative energy sector looking specifically at wind, solar and tidal/wave technologies. We began with the intention of limiting our scope to the US only, but given the global scope of the alternative energy market, and the fact that almost all the market leading companies have grown in foreign countries where the markets for this technology have been biggest, we chose to include Germany, Denmark, Spain, and China in our long-term research. The European countries represent three of the biggest markets for wind and solar technology, and are home to some of the biggest companies producing the technology[8]. China is the newest and biggest market entrant into the solar market, and could become the biggest producer of this technology over the next few years. Our plan is to follow the alternative energy market and identify the levels of openness and closedness in the areas where innovations are happening. We chose these three technologies with the expectation that we would find variations among their approaches to openness and closedness, and because the technologies represent different levels of maturity. The maturity can be measured both by the stage of development of the technology and the stage of development of the market. Wind is considered a mature technology because it is fairly well understood, and the cost of generating electricity with wind turbines is closer to the cost of conventional sources of fossil fuel generated electricity – though it is still more expensive. Solar technology is less mature and can be quite expensive, therefore the research and innovation around solar technologies is sure to play a critical role in bringing its costs down. Tidal/wave technology is relatively immature compared to wind and solar, and is mostly in the demonstration phase at this time. Only a few small projects around the world are generating consumer electricity.

These technologies are a small subset of the many AE technologies that exist, and they are all representative of energy supply technologies, meaning they are focused on bringing energy to a point of final use. There is another set of technologies called energy end-use technologies that we do not focus on. These technologies are concerned with the most efficient use of the supplied energy. Examples are home appliances, automobiles, and light bulbs.

Within our three focus technologies, wind, solar and tidal/wave, there is a variety of subset technologies.  Figure 1 provides a description of the technologies our research is focused on. These technologies are only used for electricity supply. Technologies we are not researching are solar thermal, which uses the suns energy to heat water for home and commercial use, solar heating and cooling, which uses building design to take advantage of the sun's direct heat and energy to efficiently heat and cool buildings at different times

---

[8] For instance, wind power dominates the European renewables deals tables, more so than any other major region, accounting for 60% of total European renewables deal value. Hydro accounted for the largest North American deal value in 2008, but this was almost entirely attributable to one deal. After hydro power, wind and solar power delivered the highest value deal segments, accounting for $1.4 billion and $1.3 billion of deal value respectively. http://technology4life.wordpress.com/2009/03/22/trends-in-renewable-energy-deals-in-2008-by-pwc/

of the day and during different seasons, and any wind or tidal/wave technologies that use the energy from the source for mechanical work, rather than for conversion to electricity. These excluded technologies are less common than the electricity supply technologies we are researching. We are focusing on electric supply technologies since these can have the biggest impact on reducing global carbon emissions by reducing the use of coal for power generation. Reducing the use of coal can facilitate the shift to a lower emissions plug-in hybrid vehicle market thereby reducing the world's dependence on oil.

**Figure 1**
*Technologies for electrical generation from solar, wind and tidal/wave energy*

| Alternative Energy Source | Technology Types | Technical Characteristics |
|---|---|---|
| Solar | Silicon Crystalline Photovoltaics (PV) | The most common PV panels on the market. Made from mono or polycrystalline silicon, which converts sunlight directly into direct current (DC) electricity. Conversion efficiency from sun energy to electricity is of 16% - 17%. |
| | Thin Film Photovoltaics (PV) | A less common technology in the global market, but represents 30% of the US PV market. Panels use flexible materials like copper indium diselenide or cadmium telluride and some thin silicon designs, and can be mass produced at low cost. Lower efficiencies of around 8% - 11%. |
| | Concentrating Solar Power (CSP) | A straight forward technology requiring a concentrator (mirrors, for example, that concentrate the suns rays), a receiver (a pipe or other surface that comes in contact with some material that stores the suns heat), a heat transport mechanism (a material capable of storing the sun's heat and transferring it through a heat exchanger to the water which will create steam) and a conversion system (a turbine) that drives the electrical generator. |
| Tidal/Wave | Tidal Barrage | An estuary or bay with a large natural tidal range is artificially enclosed with a barrage. Electrical energy is produced by allowing water to flow from one side of the barrage to the other through low head turbines |
| | Tidal Current | Technologies are similar to those used for wind energy conversion, that is, turbines of horizontal or vertical axes ("cross-flow" turbines) capture energy in the tidal currents. The density of water is some 850 times higher than that of air, so the power intensity in water currents is significantly higher than that in airflows, therefore water current turbines can be built much smaller than an equivalent-powered wind turbine |
| | Wave | Wave energy technologies capture the undulations of the ocean caused by wind or ocean motion and turn them into |

| | | electricity through various turbine functions. |
|---|---|---|
| Wind | Onshore | The more conventional and well understood wind turbines that range from a few hundred kW (kilowatts) in power to 2-3 MW (megawatts). |
| | Offshore | The newer and more expensive turbines built to be installed on the ocean and capable of up to 5MW of power. Must be engineered to withstand the increased forces of wind on the ocean and the rough seas. |

**Sources:** Author illustration based on information from: (Bosik 2009; Capello 2007; Capello 2008; Carlin 2004; Lemonis 2004; Luzzi & Lovegrove 2004; Perlin 2004)


*V - Technology History*

While the origins of these energy supply technologies are all based in the 1800's, the practice of using the wind, sun, and tides/waves as sources of energy for work, are much older. Wind was used to power sailboats up to 5,500 years ago, and there is evidence of windmills for mechanical work in India 2,500 years ago. Solar energy is the basis of most energy on earth, including the energy in plants from photosynthesis, solar thermal heating, the fossil remains of organic material in oil and coal, and wind which is created when air, heated by the sun, rises and cold air from another area moves into that space. Using moving water for power can be traced back to 250 BC. Wind turbines for electrical generation were first developed simultaneously in the US and Scotland around 1887. In the 1970s, NASA funded research at the Lewis Research Center in Cleveland, Ohio, to refine the design and function of electrical wind turbines. The first favorable government subsidy policy for alternative energy was introduced in the US in 1978 - The Public Utilities Regulatory Policy Act (PURPA) - which encouraged the installation of over 1400 MW of wind power capacity in California. Most of the turbines installed were built in Denmark by the leading manufacturer at that time (still the top manufacturer today), Vestas.

The solar photovoltaic (PV) effect was discovered in 1839 by Alexandre-Edmond Becquerel. He observed that when selenium was exposed to sun a small electrical current was created. Solar PV panels remained undeveloped until 1953 when the first commercial panels were manufactured at Bell Laboratories after one of the lab's scientists discovered that silicon could be used in place of selenium as a more efficient material for creating electricity. The US government took a keen interest in the technology for use in the space program, and funded PV developments for that purpose. While the history of PV panels stretches back over 50 years, it is still a very immature and expensive technology that has a small but growing market share in the US and the world. Countries like Germany, Spain, and Japan have been the market leaders in terms of PV installations.

*Figure 2*
*Top 10 Wind and Solar PV Companies*

| Wind Companies | Country (Headquarters) | Percent of market | Solar Companies | Country | Production Share |
|---|---|---|---|---|---|
| Vestas | Denmark | 28% | Sharp Electronics | Japan | 17.5% |
| Gamesa | Spain | 16% | Q-Cells | Germany | 10.0% |
| GE Wind | US | 15% | Kyocera | Japan | 7.0% |
| Enercon | Germany | 15% | Suntech Power Holdings Ltd. | China | 6.5% |
| Suzlon | India | 8% | Sanyo | Japan | 6.0% |
| Siemens | Denmark/Germany | 7% | Mitsubishi | Japan | 4.5% |
| Nordex | Germany | 3% | Motech Industries Inc. | Taiwan | 4.0% |
| RePower | Germany | 3% | Shell | UK | 3.5% |
| Acciona | Spain | 3% | SunPower | USA | 2.5% |
| Goldwind | China | 3% | First Solar | USA | 2.5% |

*Sources: Authors illustration based on information from: (Efiong & Crispin 2007; Capello 2008)*

Solar thermal furnaces that generated sufficient heat to produce steam – the basis of a CSP plant – were first developed in the eighteenth century and used in small scale applications in the US and France during the 1860's. Today, the US is seeing renewed interest in CSP plants, while the current supply of CSP generated electricity comes from a number of 80MW (megawatt) plants in Southern California, which were built in the late 1980's. CSP is a mature and very well understood technology with growing adoption in the US.

Tidal and wave technologies share a great deal in common with hydroelectric infrastructure, but new and innovative technologies are now dominating the tidal/wave industry. The first bona-fide tidal energy plant was constructed in France, at La Rance in Brittany between 1961 and 1967. It consisted of a barrage across a tidal estuary that utilized the rise and fall in sea level induced by the tides to generate electricity from hydro turbines. Today there are very few tidal/wave plants in use, and those that do exist are mostly demonstration projects. No single technology has emerged as the market favorite but over a thousand patents exist. The future promise of tidal/wave technology is great both in terms of total amounts of energy that can be generated, and the cost-competitiveness of the technologies.

## VI - About the markets

According to New Energy Finance, the clean energy sector grew to over $148 billion in 2007, up forty-one percent from 2006. Its importance is not only environmental, but also geopolitical. The goal of energy security - a deeper political issue - can be understood by the USA objectives of reducing the dependency on foreign sources of fossil fuels and controlling prices, and achieving a diversity of energy supplies. The technologies that form alternative energy – and companies that explore them - vary immensity in type, innovation cycles, maturity and techno-economic readiness.

In terms of constituencies, the presence and influence of actors vary among countries, imprinting different forms to the organization of alternative energy innovation For instance, in Japan, the government has traditionally taken a strong role in coordinating such activities through its Ministry of Economy, Trade, and Industry; while European countries have stressed and exemplified cross-country collaboration and coordination. In the US, the private sector exercises greater autonomy, even after the emphasis on public-private partnerships since the 1990s. In developing countries, such as Brazil, the government typically takes a very strong role in funding and coordinating innovation in energy, as in the biomass efforts of Petrobras. The various entities collaborate in a range of combinations, within countries and internationally, and impacts the availability of funding for R&D. For instance, the private sector accounts for the majority of expenditures for energy R&D in International Energy Agency (IEA) member countries, although governments account for a large fraction as well.

## VI.1 - Alternative Energy Policies and the Global Market

It is apparent that the market for these alternative energy technologies, and the market leading companies, have generally developed in the areas of the world with the most generous demand-pull policies, and, predictably, under governments that have prioritized the growth of alternative energy technologies. The majority of the biggest and most successful wind and solar technology companies in the world are located outside of the US, with wind manufacturers being disproportionately grouped in Germany, Spain and Denmark, and solar companies being more widely distributed between Germany, Japan, China and the US. What distinguished these countries from the US are their government's alternative energy policies. In Germany, Spain and Denmark a back-side deployment policy called a Feed-in Tariff[9] has been responsible for the rapid growth of their alternative energy technology markets, and has thus encouraged the development of many of the leading technology companies. China, on the other hand, has taken advantage of the growing market for solar energy technologies, and has funded significant R&D to create cheap and efficient solar photovoltaic cells that are being sold in foreign markets, most notably the US and Europe. There is not a significant domestic

---

[9] A Feed-in Tariff is a government statute that requires electric utilities to buy renewable energy from producers at a premium per kWh price and guarantees these prices in long-term contracts of 10-20 years. The result is a low-risk investment structure for alternative energy plants since income can be predicted with a high degree of accuracy, and income is guaranteed due to the long-term contract.

market for its solar cells, and China does not provide many government demand-pull deployment subsidies.

## VI.2. - Changes in the US Market

The 1973 oil embargo caused the US and Europe to prioritize alternative energy investment and development, providing a buffer from the volatility of supply and demand for oil. The supply-push and demand-pull policies targeting alternative energy technologies, which were initiated in the late 1970s defined the market leaders (Germany and Denmark) and those left behind (the US). Ultimately, the US was able to take a haphazard approach to alternative energy policies due to its prodigious stores of coal, oil and natural gas and political leadership that favored these industries. Now, given the national and global concerns mentioned above, the US is making a late entry into the global alternative energy market. Initially, there have been some very promising statements made by the Obama Administration. In a speech at the Brookhaven National Laboratory in late March, the new US Secretary of Energy, Steven Chu, stated that he believed it would be appropriate for the US and other countries to work collaboratively and share as much intellectual property as possible around alternative energy technologies that can help reduce global carbon emissions. Similarly, Massachusetts Congressman, Representative Edward Markey, introduced a resolution on February 13, 2009 that asked the United States to join the ***International Renewable Energy Agency (IRENA),*** a multi-national organization whose membership includes 79 countries pledging to facilitate the global growth of renewable energy through the sharing of all relevant information including renewable energy resource measures, best practices, effective financial mechanisms, and state-of-the-art technological expertise. While there is some support in the US Government for membership in IRENA, and the sharing of alternative energy technology IP with other countries, there is a strong lobby against this plan from the private sector. Companies like GE have stated that they do not intend to give their technology innovations away for free, and that doing so would undercut their ability to make profits.[10] One of the reasons for sharing this information would be to help developing countries gain access to clean alternative energy technology to help improve their living standards and grow their economies.

Additional promising news was delivered on April 27, 2009 when President Obama announced that he hopes to increase government R&D funding for new technologies, including alternative energy technologies, to over 3% of GDP, a higher percentage than the US reached at the peak of the Space Race in 1964. A financial commitment of this level will be needed as the challenges of encouraging growth in the alternative energy sector are unlike any of the US's previous technological challenges. No single alternative energy technology will be sufficient to replace conventional carbon emitting energy sources. Alternative energy will require cost-effective development of all available technologies to succeed. Direct competition with the powerful coal, natural gas and oil industries and their lobbyists will make balancing government funding difficult because the government is simultaneously and extensively subsidizing both fossil fuels and alternative energy.

---

[10] http://dotearth.blogs.nytimes.com/2009/03/26/energy-chief-seeks-global-flow-of-ideas/

At this point in our research we have not identified any commons-based models of technology development. What we have found are free renewable energy resource maps that provide measures of potential energy in particular regions on the country. The maps with this information are provided freely by the government through the National Renewable Energy Laboratory (NREL) website and consist of wind, solar, geothermal and biomass resource maps[11]. These maps are of the entire US and typically will provide detail down to 1km x 1km squares that rank the level of sun insolation in that area, the speed and consistency of wind in that area, the presence of biomass materials for harvest, or the existence of geothermal heat wells. This information is critical to developers or individuals who want to assess the viability of installing alternative energy technology in a particular location. The government believes that by providing this information for free it is encouraging the development of new alternative energy plants.

## VI.3 - Production and distribution cycle

Further research is necessary to understand how production and distribution practices incentivize openness or enclosure regarding the knowledge necessary for innovation in this sector.

In any case, the stages of Energy Technology Innovation comprise fundamental research, applied research, development, demonstration, pre-commercial and niche deployment, and widespread deployment (often also called diffusion). Technology transfer between countries is often envisioned as a part of diffusion, but it can also occur at earlier stages. Figure 4 below illustrates a schematic for observing this non-linear process specifically for energy technologies.

## VII - About policy

The main USA government policies for alternative energy aim to lower the cost of wind, solar and tidal technologies and make them cost-competitive with conventional energy sources (coal, natural gas, nuclear). These public incentives include tax credits, and state subsidy programs for wind, solar, hydro, and geothermal, biomass, biogas, and others.

Government funding policy for the research and development of alternative clean technology is provided by Department of Energy through the national energy laboratories listed under public funding. These labs offer technology transfer programs that license their patents to interested private-sector companies. There is also a Cooperative Research and Development Agreement (CRADA) for collaboration between a lab such as The National Renewable Energy Laboratory, and a partner company. However, IP ownership and sharing of knowledge is a deal-base negotiation.[12]

---

[11] http://www.nrel.gov/renewable_resources/
[12] Two are the main models for the CRADA: "Shared-resources" which means the research is funded by the government and is part of ongoing research at NREL. In this case no funds change hands and "Funds-in"

Further research is necessary to understand how these practices incentivize openness or enclosure.


## VIII. Funding

Federal support for energy R&D has fallen by more than half since a high point in 1980, and private-sector energy R&D has similarly fallen. These levels of expenditure compare poorly to other major federal R&D efforts that met challenges of similar magnitude: the Manhattan Project, the Apollo Project, the Carter-Reagan defense buildup, and the doubling of the budget of the National Institutes of Health. Advances in energy technology will not occur on the scale required without significantly increased investment by both government and business. In contrast, President Obama announced on April 27, 2009 that he hopes to raise the level of R&D funding in the US to over 3% of GDP, a higher level than was reached during the space race. One of the main research areas for this funding would be alternative energy.


## VIII.1. - Public R&D Funding

As mentioned in the introduction, the US reduced its public R&D funding throughout the 80s, 90s and into the new millennium, and only in 2006 did R&D funds start to increase. Most of these funds are being given to the 17 U.S. Department of Energy laboratories, which have historically been an ineffective model for alternative energy development and commercialization. The main reason for this ineffectiveness is that most of the labs do weapons research, which is developed for one client, the U.S. Government, and as a result the lab system as a whole, lacks the private sector business acumen to launch the technologies from initial innovation through demonstration across the "valley of death" and into commercialization. See figure 3 below for a list of the laboratories. Another portion of public R&D funding goes to universities. The top research universities have very well developed technology transfer offices that advise students and professors on their patenting policies, and help them figure out the best approach to moving their innovations from the university into the market. On the other hand, many universities are not as savvy in this respect, and some innovations are lost due to administrative hurdles. The Bayh-Dole Act, adopted in 1980, created a reverse presumption of ownership in favor of the government for small businesses, universities, and non-profits that accepted government funding for their research. Despite this presumption, the act created an opportunity to pursue intellectual property protection and commercialization of the invention.

which means the partner will pay for all or part of the research, but NREL does not provide the partner with any funds.
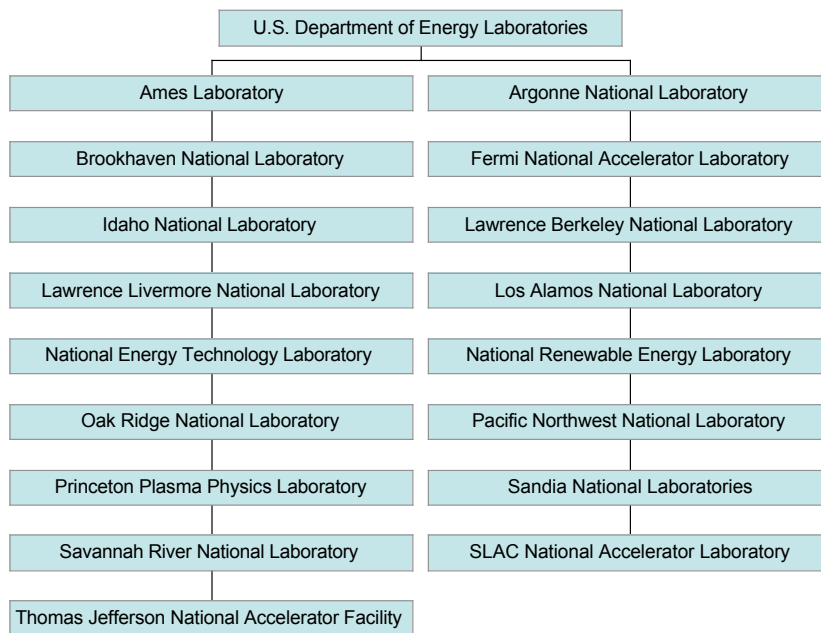
```
┌─────────────────────────────────────────────┐
│        U.S. Department of Energy Laboratories │
└─────────────────────────────────────────────┘

┌──────────────────────────┐      ┌──────────────────────────────┐
│      Ames Laboratory      │      │  Argonne National Laboratory  │
└──────────────────────────┘      └──────────────────────────────┘

┌──────────────────────────┐      ┌──────────────────────────────┐
│ Brookhaven National       │      │ Fermi National Accelerator    │
│ Laboratory                │      │ Laboratory                    │
└──────────────────────────┘      └──────────────────────────────┘

┌──────────────────────────┐      ┌──────────────────────────────┐
│  Idaho National Laboratory │     │ Lawrence Berkeley National     │
│                           │      │ Laboratory                    │
└──────────────────────────┘      └──────────────────────────────┘

┌──────────────────────────┐      ┌──────────────────────────────┐
│ Lawrence Livermore        │      │ Los Alamos National           │
│ National Laboratory       │      │ Laboratory                    │
└──────────────────────────┘      └──────────────────────────────┘

┌──────────────────────────┐      ┌──────────────────────────────┐
│ National Energy Technology │     │ National Renewable Energy      │
│ Laboratory                │      │ Laboratory                    │
└──────────────────────────┘      └──────────────────────────────┘

┌──────────────────────────┐      ┌──────────────────────────────┐
│ Oak Ridge National        │      │ Pacific Northwest National     │
│ Laboratory                │      │ Laboratory                    │
└──────────────────────────┘      └──────────────────────────────┘

┌──────────────────────────┐      ┌──────────────────────────────┐
│ Princeton Plasma Physics  │      │ Sandia National Laboratories   │
│ Laboratory                │      │                               │
└──────────────────────────┘      └──────────────────────────────┘

┌──────────────────────────┐      ┌──────────────────────────────┐
│ Savannah River National   │      │ SLAC National Accelerator      │
│ Laboratory                │      │ Laboratory                    │
└──────────────────────────┘      └──────────────────────────────┘

┌──────────────────────────────────┐
│ Thomas Jefferson National         │
│ Accelerator Facility              │
└──────────────────────────────────┘
```

*Figure 3*
U.S. Department of Energy Laboratories
Source: http://www.energy.gov/

### VIII.2 - Private R&D Funding

Experts believe that the private sector provides a larger portion of the R&D funding for alternative energy technologies, but due to the proprietary nature of the funding information within private companies, it is very hard to track down how much money has been spent and on what technologies.[13] The access to information is limited starting from early stage angel investing and continuing through mature venture capital contributions. We're looking into opportunities to gather this information either through interviews, or perhaps surveys as part of the next steps in our research.

The following figure represents the value chain for the alternative energy market, which

---

[13] The Kennedy School at Harvard University - Energy Research, Development, Demonstration & Deployment Policy, Energy Technology Innovation Policy, coordinated by Laura Diaz Anadon.

is more complicated than many other business sectors.

*Figure 4*
*Actors and their interactions in the alternative energy market*
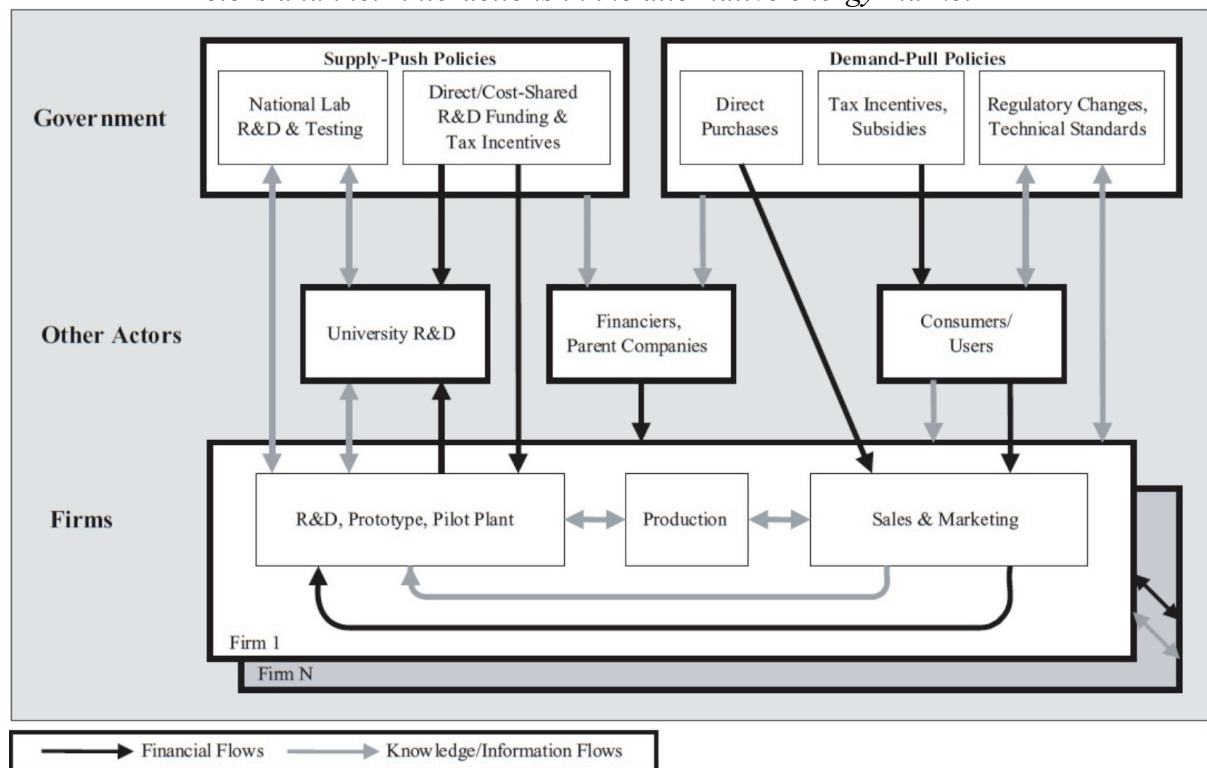


Source: (Gallagher et al. 2006)

Figures 5 and 6 below show comparisons of OECD countries' (Denmark, Germany, Spain and the US) government investments in research, development & demonstration (RD&D) from 1974 – 2007. Figure 3 is a comparison of the various countries' investments in the specific technologies we are researching – solar, wind, and ocean energy.[14] Figure 4 is a comparison of the various countries' investment in these

---

[14] Ocean energy includes ocean thermal energy as well as tidal/wave, and though we are not researching ocean thermal technology, the International Energy Agency database does not provide a breakdown between the investments in these three technologies.

technologies as a percentage of the total amount invested in all energy RD&D, including conventional sources. When these energy investment numbers are compared to the total gross domestic product (GDP) of each of these countries, the percentage of energy investment is minute – well under 1% for all four countries. We chose not to graph those results due to the very small variations in the numbers. The most obvious spike in alternative energy investment is the increase in US RD&D for solar, wind, and ocean energies between 1974 and 1980. This can be attributed to post oil embargo concerns about developing alternative energy technologies (discussed below), and President Jimmy Carter's focus on alternative energy and environmental stewardship. 1980 marked the year that President Ronald Reagan took office. The resulting drop in RD&D investment can be attributed to his conservative politics and his favoritism towards the oil, gas and coal industries.
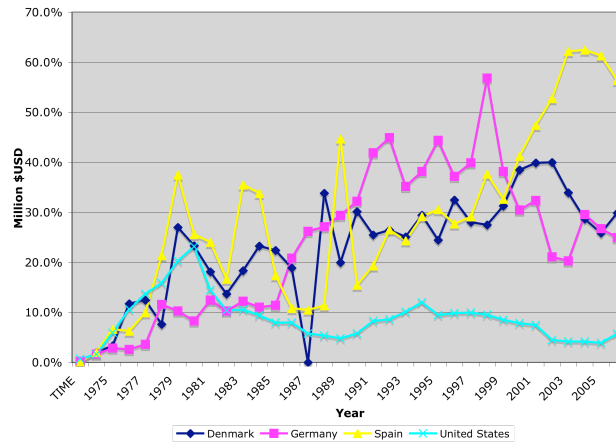
*Figure 5*

**Solar, Wind & Ocean RD&D Budgets by Country(1974 - 2007)**
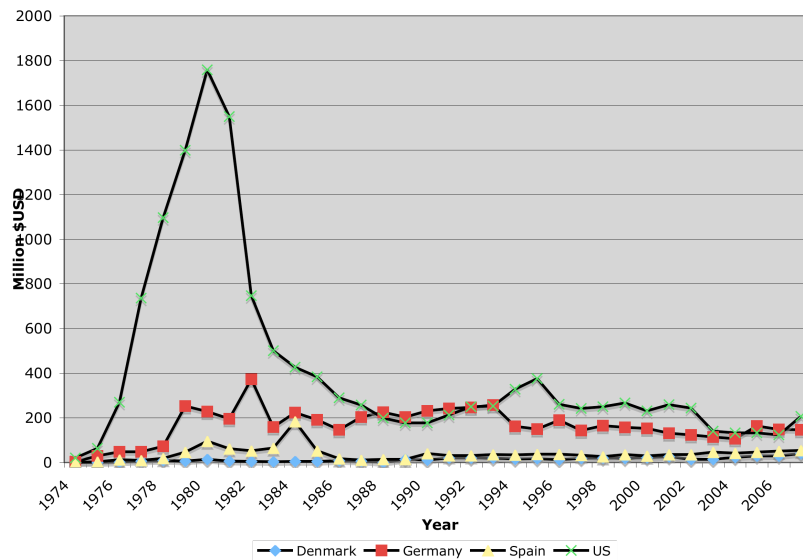


Source: Authors illustration based on information from: IEA 2009

*Figure 6*

**Solar, Wind & Ocean RD&D Budget as % of Total Energy RD&D**



Source: Authors illustration based on information from IEA 2009

## IX - Cooperation Mapping Quadrants

Unlike the other sectors we chose to analyze, we could not identify any clear commons-based models in the Alternative Energy at this point of the research. In addition, since the discussion regarding Intellectual Property in this field is emerging, a political map of actors is also less clear. However, with the appearance of organizations such as I.D.E.A. and IRENA, and the reorganization of public-funded laboratories by the new administration in the USA, we expect to begin mapping AE actors and clouds into the quadrants tool in the second phase of our research.

In this sense, many questions remain open. Is there cooperative work happening at such a small scale that we could not find it? Is there something blocking cooperation? How could cooperation be beneficial in this field? What may foster cooperation in this field? What will be the impact of the Obama administration's push to share technology versus private pushes towards regulation?

## X - Trends in the field

Since our research focus is related to how knowledge flows in order to spur innovation in the AE field, and if commons-based models are emerging, the trends we are interested in relate to how practices around business models and use of Intellectual Property evolve in a certain sector. Thus, there are multiple current trends that call our attention in the Alternative Energy sector: patent growth; new organizations forming in order to foster cooperation (and their IP strategies); new organizations forming to promote regulation (in this sense, we must pay close attention to associations mentioned above such as I.D.E.A. and IRENA) and, finally, what the barriers that characterize this sector and their impact to foster cooperation.

*X.1 - Patent Growth*

Patent growth can be interpreted as an indicator of market growth, and we have found that the last six years have showed a steady increase in alternative energy patents, which consequently coincides with the adoption of more demand-pull policies, and, since 2006, the increases in supply-push policy funding. Figure 5 below shows this increase.

*Figure 7*



Clean Energy Patent Growth Index by Year
2002 - 2008
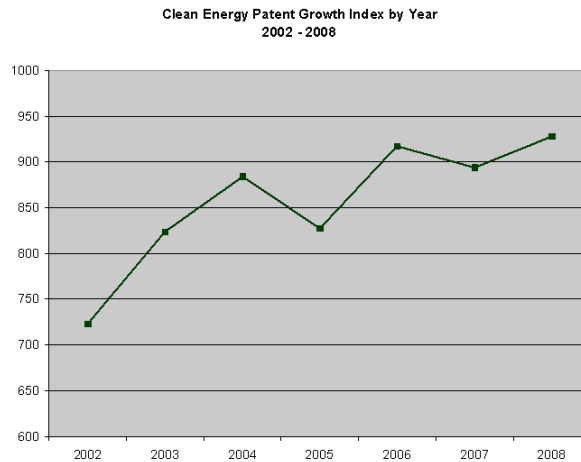
Source: (CEPGI 2009, Published by Heslin Rothenberg Farley & Mesiti P.C.,
http://cepgi.typepad.com/heslin_rothenberg_farley_/)

*X.2 - Barriers*

There will be a number of barriers that make the adoption of alternative energy technologies quite difficult. As mentioned above these disruptive energy supply technologies will have to battle the incumbent conventional energy supply technologies and their powerful lobbyists. It will also take strong political will on the part of the President, Congress, and the voters, to maintain the supply-push and demand-pull policies that support these technologies and the market for them. Permitting creates another barrier, especially in the cases of tidal/wave technologies, and offshore wind, which is built in a region whose use is already contested by many major players (ie. Fishing, Offshore Drilling, and Shipping) and whose ecosystem is already delicate. Onshore wind and CSP plants face various permitting procedures on land, and are often the targets of NIMBY – Not in My Backyard – complaints from local residents who object to the visual impact, environmental impact, or noise in the case of wind turbines.[*] Solar PV is the one technology that suffers far fewer NIMBY concerns due to its integration into the built environment rather than the natural environment.

Wind energy, solar energy and wave energy are subject to intermittency due to the unpredictable nature of the source. This means that wind, solar and wave energies can't

---

[*] NIMBY concerns are also very common in conventional energy supply technologies since residents are never happy to have a new oil refinery, or oil, coal, gas or nuclear plant in their neighborhood.

be relied on at all times, just the times when the wind is blowing or the sun is shining or the waves are rolling. Until adequate and affordable energy storage solutions are created – like batteries or compressed wind storage – these energy technologies can't be used for baseload electricity supply, which is any electricity that is available 24/7 to meet the real-time demand of consumers.

The cost of grid interconnection creates another barrier for new alternative energy technologies. Running new transmission lines from the rural areas where onshore wind and CSP plants are typically located can be complicated and expensive. Permitting procedures for the siting of the lines, which are similar to the permitting issues mentioned above, can take a long time and cost a great deal in consultant and lawyer fees. NIMBY concerns from people located close to the new lines, can lead to political battles and more legal battles, especially in cases when proposals include line passages over private land. The distances that must be covered to interconnect rural plants to major electricity transmission corridors, exacerbate the transmission issue even more. All of these same issues are present for offshore wind and tidal/wave plants, with the added complications of offshore permitting procedures as described above. Solar PV is the one technology that avoids many of these issues when it is installed in the built environment. Most PV installations connect to the electrical grid through a property owner's electrical distribution system (the system of low kilovolt lines that carry electricity around neighborhoods and into buildings). Excess electricity produced by these PV installations takes a backwards route from the distribution system, back through a substation where its voltage is increased for transport to the high kilovolt transmission lines.

At the current time, alternative energies are facing a difficult funding challenge due to the global economic downturn. Venture capital funding has been reduced while the credit markets have dried up leaving very little access to capital for new projects. Another contributing factor are the credit equity markets, which are tied to the tax credits offered for alternative energy projects in the US. Since most alternative energy developers do not have a high enough tax liability to take advantage of the tax credits offered by the government as an incentive to build an alternative energy plant, the developers sell the tax credits to large banks who apply the tax credits to their own tax liability, and in the process, become a part investor in an alternative energy project. Currently, the banks that would have purchased the tax credits don't have the capital to do so, nor do they have the tax liability that would necessitate a tax credit, due to their shrinking profits. The result is that there is a lack of funding for alternative energy projects throughout the world, and in the US.

## XI - Next steps

In the next steps of our research we plan to:

1. Look more closely at the international history of the technological innovations that led to the current alternative energy market landscape.
2. Put more emphasis on case studies of certain companies in the alternative energy

sector that may exemplify the open or closed models that the research is based on.

3.  Possibly add surveys for specific topics of importance like information about private R&D funding in collaboration with Kennedy School.
4.  Follow the activities and conduct a case study of IRENA as a global example of collaboration on alternative energy technology.
5.  Follow the activities and conduct a case study of I.D.E.A. as a local example of political barrier to cooperation on alternative energy technology.
6.  Conduct more research on the US DOE National Laboratories to understand how they develop new technologies, how they are patented and licensed and their funding structure for alternative energies. And how their practices impact in the emergence of closedness or openness arrangement.
7.  Research the IP positions of alternative energy industry associations and their activities around the Patent reform in the USA.

# ICP PROGRESS REPORT ANNEX 4

## Educational Materials Sector

## *I- Introduction*

The field of educational materials (EM) refers to a subset of the book, games, Internet, and software publishing industries that is focused on providing resources to a variety of educational market segments. For instance, PricewaterhouseCoopers characterizes the EM sector as divided into digital and non-digital solutions. At the K-12 educational level, digital solutions include a range of technologies used to enhance the delivery and the administration of K-12 education, including data management systems, web-based course and assessment materials, and online tutoring and professional development—however, we will only focus on those digital solutions products that have specific educational purposes and where knowledge is embedded in a form that can be enclosed by some form of intellectual property. Regarding non-digital solutions, we include textbooks, course packs and other supplementary materials, and various educative toys and games.

Actors providing these materials are private companies such as publishers controlling the textbook and complementary materials markets; global media companies focused on the family-based market, such as the Discovery Channel; public institutions, such as National Public Radio; universities and their presses, providing both closed and open educational materials; and independent organizations and associations comprising educators and interested individuals wanting to contribute to the open educational resources (OER) movement.

## *II – Research Focus*

The EM field represents a unique case for studying the trends in access to and regulation of a commons because of its long history and its established ideological clusters. Of particular interest to our study is the prominence of the OER movement. To quote the Center for Educational Research and Innovation, "learning resources are often considered key intellectual property in a competitive higher education world, more and more institutions and individuals are sharing digital learning resources over the Internet openly and without cost, as open educational resources". The definition of OER currently most often used is "digitized materials offered freely and openly for educators, students and self-learners to use and reuse for teaching, learning and research". OER involves educational content, the software tools to develop, use and distribute that content, and implementation resources such as open licenses. Excellent examples of OER are the MIT OpenCourseWare project started in 2002 or the Connexions Platform, both in the USA, the Projeto Folhas or the Professor Portal in Brazil, or the Siyavula Project in South Africa.

Spurred in part by the increasing momentum of the OER movement, EM producers, consumers, and various regulatory and advocatory bodies are each responding to and evolving with digital technology, and thereby creating new market demands and offerings. While these trends are complicated in their economics, politics, and other social factors/barriers, the consolidation of the field's largest producers (specifically publishers of the still most important form of EM: the textbook) allows for our descriptive research of the field and ongoing innovations to be manageable and ultimately mappable using our Quadrants tool. Furthermore, the different educational levels offers us glimpses at how and why innovation trends differ, considering that the Higher Education level seems to be more concerned with the price of EM while the K-12 level is more concerned with quality and diversity of available resources.

### III - The Intellectual Property Factor

Copyright represents the most significant intellectual property tool involved in this field due to the textual nature of its outputs. Patents are also involved when factoring in the educational software segment; however, educational software represents a significant amount of copyrightable material as well. The use of strategies focused on copyright is one of the fundamental points on which ideological differences can be identified between private companies, the set of universities and independent organizations interested in OER, and the advocatory associations that represent each. Similar to the issues facing the newspaper industry, and by analogy the film and music industries, the protection afforded by copyright has become uncertain as EM goes digital and the ability to monetize digital distribution still presents a challenge and potential barrier to innovation.

### IV – Research focus: Our Early Focus on Textbooks

Although our study is not limited to textbooks, particularly when one considers how digital expressions of educational content have begun to blur the definition of a "textbook", the prominence of this form of EM acts as an important gateway to understand the rest of the EM field. The price of and access to quality textbooks, particularly at the higher education level, has been a recent highly controversial issue. The Student Public Interest Research Groups (PIRGs) have brought the issue widespread attention through their Make Textbooks Affordable campaign, inaugurated in 2004 when they published a research report entitled Rip-off 101: How the Current Practices of the Publishing Industry Drive up the Cost of College Textbooks. As noted in trade magazines like Publishers Weekly, the PIRG study criticizes the industry on the following points:

- textbook prices are rising "at more than four times the inflation rate for all finished goods";
- the most popular texts have new editions published every three years;
- new editions are priced 12% higher than the editions they are replacing, "almost twice the rate of inflation";
- bundled texts are on average 10% more expensive than unbundled counterparts;

- 55% of bundled textbooks are not available unbundled;
- the average textbook costs 20% more in the U.S. than U.K. and some are dramatically more expensive.

Subsequent government-funded studies and policy movements have created a proliferation of research on the textbook market that unites the ideological questions of access and regulation. Simultaneously, technological affordances have resulted in alternative models of EM production and delivery gaining ground in both the K-12 and higher education levels.

## V - Technology History

Michael Watt has traced the history of US textbook publishing to the 1880s and ascribes the emergence of the textbook to "greater uniformity in local education systems resulting from immigration and industrialization". For Watt, "the development of modern practices in textbook publishing in the USA was concomitant with the rise of mass education, characterized by graded organization of formal schooling into classes."

Established publishers quickly took control of and stabilized the new textbook market; and before World War I the American Book Company had formed a monopoly. New publishers proliferated after the war, and in 1931 the National Society for the Study of Education pushed for a standardized culture of publishers soliciting manuscripts and judging the innovative merit of each work and the competency of the respective authors.

During the 1950s and 60s, textbook publishing became more competitive but remained largely professional with companies led by founder-editors. The rise of the role of official state adoption, particularly of textbooks in each K-12 subject, coincided with this moment representing a new and important market force that heavily contributed to the sector structure we observe even today.

In the 1970s, what sociologists Patricia Thornton and William Ocasio call the "market logic" began to pervade the industry as investors became interested in the market potential of publishing houses. Many private companies went public and were purchased by investment companies, merged with other publishers, or similarly acquired. For instance, founder-editors, practitioners of an "editorial logic" focusing on reputation, were replaced by profit-maximizing chief executives.

The market was further consolidated through the 1980s and 90s. K-12 business strategies for publishers focused largely on widespread state adoption of their textbooks. Leading textbook publishers with longstanding relationships at state and local levels began to include CDs and DVDs with their textbooks to deliver modular content, and many are now acquiring technologies that add value by incorporating assessment and analytical capabilities into instructional materials. Similarly, the strategy for Higher Education has moved toward bundling supplementary materials to cover all learning styles and satisfy a desire for multimedia components. The one-stop-shop strategy (including horizontal

growth and product differentiation movements) in addition to resource modularity is becoming routine among the incumbents of the EM sector. However, the marketing approach continues to involve sales representatives approaching lecturers to individually adopt textbooks for their courses, particularly introductory, obligatory courses with large student enrollments.

Further growth in the demand for digital solutions has been caused by the ongoing impact of the No Child Left Behind Act, improving IT infrastructure in schools, and the growing number of tech-savvy students and teachers. In this market, acquisitions and mergers focusing on market penetration and product diversification seem to be the rule. Examples of this trend are Pearson's 2006 and 2007 acquisitions of eCollege, Effective Education Technologies, PowerSchool and Chancery (announced May 2006); McGraw-Hill's purchase of Turnleaf Solutions (announced in 2005), now part of The Grow Network; and Houghton Mifflin Riverdeep's (HMR) purchase of Achievement Technologies.

However, PricewaterhouseCoopers identified a number of niche players focused on software development who have emerged alongside a "variety of small entities, many with roots in academia, [...] offering open-source instructional management systems to financially strapped school districts", as well as OERs. In addition, larger software/communications companies like Intel and Verizon are starting to offer free solutions through outreach programs in order to create goodwill and gain the opportunity to sell proprietary solutions.


## VI - About the EM Markets

The 2006 U.S. Census Data (U.S. Census Bureau, Company Statistics Division & Bowan 2008) presented 3,042 firms, with 83,504 employees and annual payroll of $4,993,924 in the book publishing industry, of which textbook firms are part.

Specifically, the market share for new textbooks (that is, the $4.2 billion segment) is highly consolidated, with 6 publishers holding about 85% of all sales dollars (Pearson, Thomson, McGraw Hill, John Wiley, Houghton Mifflin, and St. Martin's/Von Holtzbrinck). Although college publishing remains highly profitable for the large players, with reported EBITDA in some instances as high as 30%, growth has stalled, due in large part to the rise of the used book business, which represents the key strategic issue in the industry today.

In terms of sales, actors in the educational materials sector were responsible for:
- Professional and scholarly materials: $4.1 billion (176 million units)
- University press materials: $450 million (24 million units)
- Elhi (elementary and high school texts): $4.7 billion (178 million units)
- College textbooks (all levels): $3.9 billion (67 million units)

## VI.1 - Production and Distribution Cycle

Lisa Shamchuk defines 15 basic steps in the production and distribution cycle of textbooks:

1. Author submits proposal to publisher
2. Publisher conducts a market review
3. A contract is established
4. Editorial team + author develop manuscript
5. Manuscript is reviewed
6. Author approved revisions
7. Cover is designed
8. Supplemental materials are prepared
9. Marketing strategy is developed
10. Index is prepared
11. Book is sent to the printer
12. Book is promoted by sales representative
13. Adoptions are secured
14. Books are shipped to the bookseller
15. Profits are distributed.

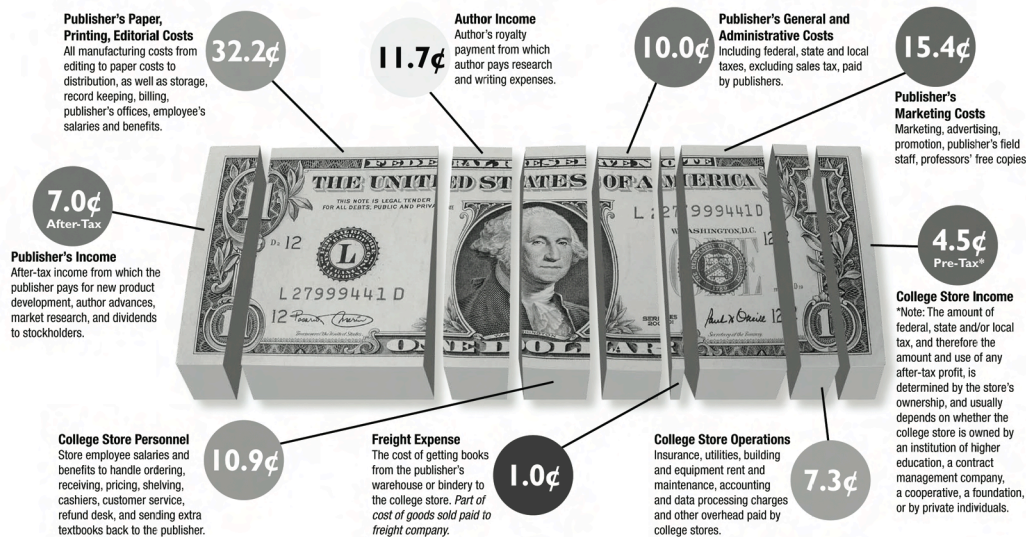The GAO report summarizes the stages in the following picture:
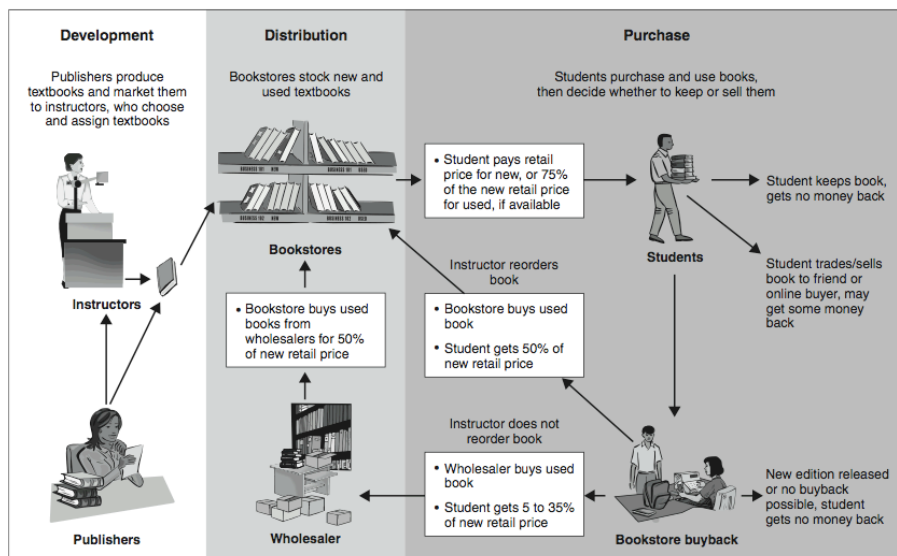


The Typical Life Cycle of a College Textbook
Source: GAO Report[15]

---

[15] College Textbooks: Enhanced Offerings Appear to Drive Recent Price Increases. Available at: http://www.gao.gov/products/GAO-05-806

## VI.2 - Costs of Production

Despite the lack of transparency in terms of textbook production costs, when analyzing financial reports of publishing industries in this sector, the National Association of College Stores created the following schematic to illustrate (and defend) the sources (as percentages of each dollar) of contemporary textbook costs:



# Where the New **Textbook Dollar** Goes*...

**Publisher's Paper, Printing, Editorial Costs** — **32.2¢** — All manufacturing costs from editing to paper costs to distribution, as well as storage, record keeping, billing, publisher's offices, employee's salaries and benefits.

**Author Income** — **11.7¢** — Author's royalty payment from which author pays research and writing expenses.

**Publisher's General and Administrative Costs** — **10.0¢** — Including federal, state and local taxes, excluding sales tax, paid by publishers.

**15.4¢** — **Publisher's Marketing Costs** — Marketing, advertising, promotion, publisher's field staff, professors' free copies.

**7.0¢ After-Tax** — **Publisher's Income** — After-tax income from which the publisher pays for new product development, author advances, market research, and dividends to stockholders.

**4.5¢ Pre-Tax\*** — **College Store Income** — *Note: The amount of federal, state and/or local tax, and therefore the amount and use of any after-tax profit, is determined by the store's ownership, and usually depends on whether the college store is owned by an institution of higher education, a contract management company, a cooperative, a foundation, or by private individuals.

**College Store Personnel** — **10.9¢** — Store employee salaries and benefits to handle ordering, receiving, pricing, shelving, cashiers, customer service, refund desk, and sending extra textbooks back to the publisher.

**Freight Expense** — **1.0¢** — The cost of getting books from the publisher's warehouse or bindery to the college store. *Part of cost of goods sold paid to freight company.*

**College Store Operations** — **7.3¢** — Insurance, utilities, building and equipment rent and maintenance, accounting and data processing charges and other overhead paid by college stores.

*College store numbers are averages and reflect the most current data gathered by the National Association of College Stores. Publisher numbers are estimates based on data provided by the Association of American Publishers.

© 2008 by the National Association of College Stores
FD.023.04.08

Compliments of the NACS **Foundation** IMAGINE TOMORROW

National Association of College Stores
CONNECT | GROW | SUCCEED

However, the textbook is only one form of EM, and not the only consideration in terms of cost affecting a publisher's success. And with the importance of state adoption at the K-12 level (something we will discuss again later), Gilbert Sewall argues that a nationally competitive company must be "capital intensive" and "full service", meaning "it must offer study guides, workbooks, and technology, along with discounts, premiums, and an array of teacher enticements"; and in states like California and Texas, "Spanish versions of texts, as well as teachers' editions, binders, and answer keys may determine which books are adopted".

## VI.3. - The K-12 Level Market

The following chart using 2008 data from Simba Information illustrates the consolidation of the traditional textbook market at the K-12 level, where just four publishers control over 90% of the market:

**2008 K-12 Textbook Publishers' Market Shares**
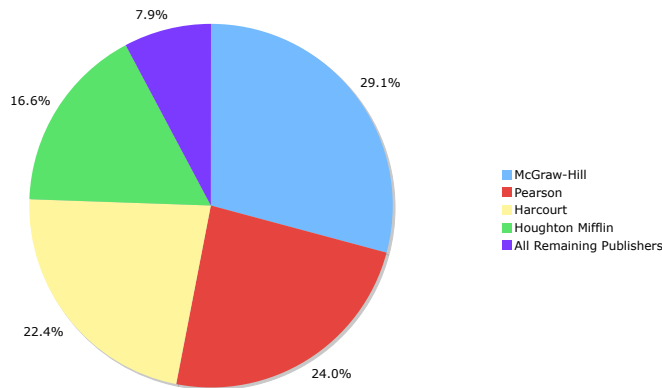


Beyond the dominance of the traditional printed textbook, some commons-based models have emerged at this level and penetrated the state approval board system. The CK-12 Foundation facilitates the peer-production of open-content "Flexbooks", and recently was contracted to produce a physics Flexbook for the Commonwealth of Virginia. In California, the Governor announced in May 2009 an initiative to produce an open-source textbook for science and math classes across the state, in order to face drastic cuts in the state budget for education.

## VI.3- The Higher Education Level Market

The following chart using 2007 data from Simba Information illustrates the consolidation of the traditional textbook market at the Higher Education level, where six publishers control over 90% of the market:

**2007 College Textbook Publishers' Market Shares**
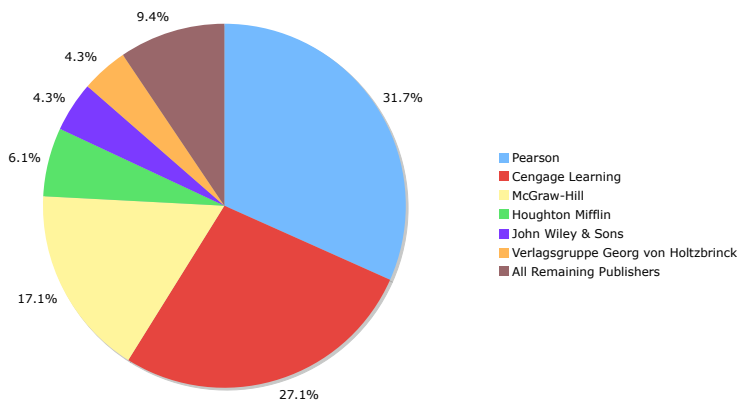
Since 2007, the Higher Education textbook market has further consolidated with Cengage Learning's acquisition of Houghton Mifflin's College Division.

Similar to the K-12 market, commons-based and alternative business models have emerged alongside the dominant textbook publishing model. Connexions is a Rice University based OER project to create educational "modules" that can be individually produced by many contributors and then brought together in textbook-like fashion for specific courses. There has been widespread praise for Connexions' earliest material on the subject of digital signal processing and also math, physics and music. In the for-profit side, Flat World Knowledge produces textbooks by single authors, available under Creative Commons BY-NC-SA licenses, which can then be "re-mixed" by students and professors for specific educational purposes, such as versioning. Another model using Creative Commons BY-NC licenses is Bloomsbury Academic, a new imprint of the Bloomsbury Publishing Group. Both Flat World and Bloomsbury offer free versions of their books with optional print-on-demand and supplementary materials purchasing. Connexions, looking for long-term sustainability, also offers print-on-demand purchasing and international partnerships to improve the use of their platform.


*VII - About the Policy: National and State Textbook Policy Trends*

Since the Student PIRG 2004 study, a number of bills have been considered at the state and national level to both examine and act on the issue of textbook prices. More recently, discussions around open the resources subsidized with taxpayer money are also in vogue.

For instance, at the state level, legislation has largely focused on nine ideas:

- sales tax exemption or tax credits for textbooks;
- requiring faculty to consider costs of educational materials;
- regulating educational materials that have particularly limited re-usability;
- requiring publishers and/or bookstores to offer unbundled textbooks in addition to bundles for particular courses;
- requiring publishers to disclose info on textbooks' wholesale prices and revision histories;
- reducing sales of new textbooks by publishers;
- recommending that institutions explore alternative textbook sources or otherwise innovate to reduce costs of educational materials (e.g. textbook rental programs);
- regulating textbook prices in public institutions
- commissioning studies and reports to investigate high cost of textbooks;
- requiring schools/bookstores to actively promote textbook buyback programs.

At least three pieces of legislation involving these ideas passed in the states of California (AB 2477 in 2004), New York (AB1214 in 2005), and Colorado (HB1024 in 2006).

At the national level two major proposals have been proposed: the College Textbook Affordability Act of 2007 (S. 945), sponsored by Sen. Dick Durbin, and the Learning

Opportunities With Creation of Open Source Textbooks (LOW COST) Act of 2009 (H.R. 1464), sponsored by Rep. Bill Foster. The former did not pass; however, the latter has been referred to committee. Foster's proposal would require certain federal agencies to collaborate and develop freely available open source educational materials for college-level math and science subjects.

Similar to Foster's proposed legislation focusing on higher education, California Gov. Arnold Schwarzenegger launched an initiative in May 2009 to "make California the first state in the nation to offer schools free, open-source digital textbooks for high school students". The initiative was a response to California's recent budget crisis, and the hope is to cut costs on expensive math and science textbooks across the K-12 public school system. This action comes along with many that try to improve Californian students' performance on national and international math and science exams[16].

## VIII - Funding

Publishing industries at the K-12 level fund their activities through the adoption process by which states purchase books, creating a small and unevenly distributed market. The twenty-one states that currently have state-wide adoption policies are mainly in the South and West and are dominated by California, Texas, and Florida, which account for as much as a third of the nation's $4.3 billion K-12 textbook market. Few "elhi" (elementary + high school) textbook publishers can afford to spend millions of dollars developing a textbook series and not have it adopted in these high-volume states.

However, an evolving issue – very similar to what has been seen in the music industry – is that a combination of cost-cutting pressures and changes in technology and user-behaviour practices (e.g.: peer-to-peer textbooks file sharing; used textbooks; course packs; photocopies; among others) are becoming a threat to incumbents in this sector.

Additionally, OER projects are, in the majority of cases, funded by private non-profit foundation funders, such as Hewlett[17], and to a lesser extent Open Society Institute and Ford Foundation. Some projects also incorporate OER practices as part of the professor's activities, with the MIT OpenCourseWare a prominent example. Currently, one of the main discussions in this area is the long-term sustainability of OER projects.

## IX - Quadrants Mapping

We have plotted our descriptive research of the major actors, market forces, and outputs of the Higher Education textbook market onto our Quadrants tool; our most recent

---

[16] In this sense, it worth pointing out that the "Adoption" States – such as California, Texas and Florida – have continuously performed worst than the "Open Territory" States. See more at: https://cyber.law.harvard.edu/commonsbasedresearch/Overall_Picture_of_the_EM-K12_field#Demand_Structure_in_K-12

[17] Experts estimate that this organization already invested $80 million in OER projects

mapping of actors with regard to representative types of commons is shown below. Even with crucial political-economy differences between the textbooks in K-12 and textbooks in Higher education, the final picture of these fields are very similar when regarding the strategies used by incumbents and challengers.
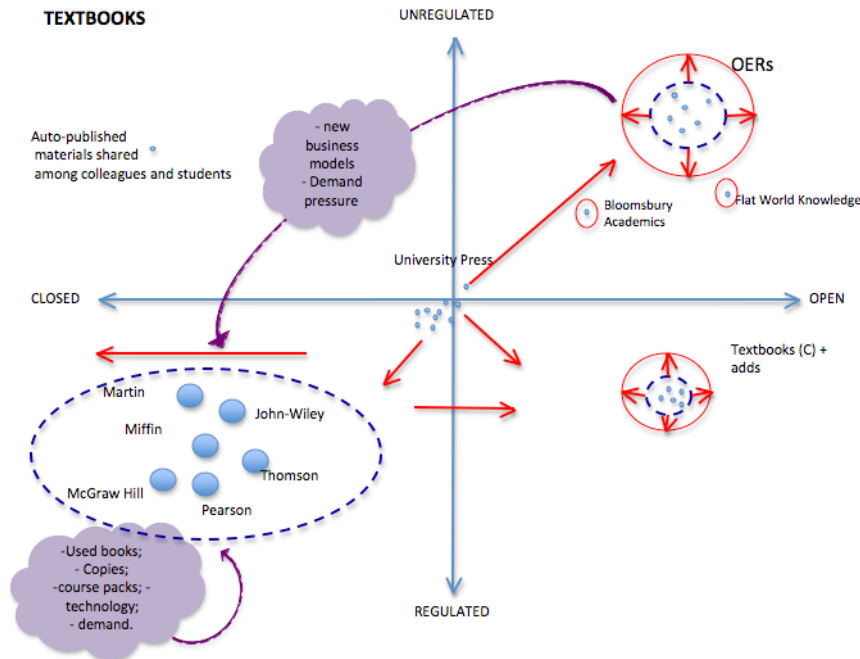


## IX.1 - Closed versus Open in the EM Field

The closed versus open axis represents the level of access to participation in the textbook production. The business models of the traditional publishers are vertical in their use of individual authors or small teams of co-authors in producing textbooks. Much of the scholarly works published by university presses are on the closed side of participation as well. On the opposite side of the participation axis, the alternative model of peer-produced and open-source textbooks sponsored by advertising and OER projects allow for open access to the production process. However, some models we could identify on textbooks based on advertisement are "free as in beer," and not "as in freedom" - representing just a distribution model for e-books, which, in general, have a printed version available for a higher price.  This relates to the placement of such models in a different part of the quadrant as they retain more regulations than models where rights are provided to users in addition to no-cost content.

## IX.2 - Regulated versus Unregulated in the EM Field

The regulated versus unregulated axis represents the level of regulation of usage embodied by intellectual property tools employed in the licensing of actors' outputs and also the propensity of the actors enforce their intellectual property rights. OER projects generally use open licensing such as Creative Commons or the GNU Free Documentation

License, which has a limited set of regulations regarding the conditions of use of resources — and in most cases enforces different levels of openness. In contrast, traditional publishers and the textbooks sponsored by advertising use copyright in its more traditional sense to severely restricts copying, distribution, and other uses not specifically allowed by the publisher in advance. These incumbent actors also enforces their rights through a series of political, legal and judicial strategies, from which suits against Universities use of course packs are a clear example.

## IX.3 - Defining the 'Clouds' in EM

The clouds represent actors that do not produce textbooks but affect the sector and can incline actors toward certain types of business strategies, which can be identified either toward closedness or openness. Most fundamental are consumers' (students'/professors') demands for quality textbooks from established publishers or for OER. Legislative bodies, as discussed earlier, can affect the market by requiring that constituent schools explore less costly forms of EM that might exemplify models that are less regulated and/or more open to participation. Professional associations can also act as clouds by standardizing intellectual property positions for member actors and advocating for a particular form of EM (like the traditional textbook). Finally, pressure from actors with different business models, such as the ones based on used textbooks, are crucial in understanding changes incumbents adopt in order to keep their dominant position in the market over time. Shorter cycles of textbooks versions and the offer of sale of textbook chapters are a clear example of a reaction.

In the case of the K-12 level, State textbook Adoption Boards are major clouds. The textbook adoption process, in place in at least twenty-one states, is a state-wide process where a central textbook committee or the state department of education review textbooks according to state guidelines and mandate specific textbooks and educational materials that all public schools in a certain state must use, or lists of approved textbooks and educational materials that these schools must choose from. Some of these states allow schools to buy other materials with non-state money. The states that do not hold adoption processes are called 'Open States'.

Texas, California, and Florida represent the largest markets for textbooks governed by state adoption boards, and publishers try to produce textbooks to specifically gain approval of these boards, which can guarantee access to key public school markets. According to the Thomas B. Fordham Institute, few publishers "can afford to spend millions of dollars developing a textbook series and not have it adopted in these high-volume states." The persistence of this trend may significantly slow innovation toward other types of commons models because of specific standards in place for these textbooks and the continuing profitability for major publishers touting approved textbooks. This barrier was already identified as a challenge for the recent California open textbooks effort.

*X - Trends in the Field*

The effects of the clouds and decisions made by major actors determine the economic, political, and social forces that move textbook production along the axes. Demand for digital delivery of EM has expanded the influence of OER, and led to major publishers and university presses exploring more open access and less regulated business models. This has also created gaps in the market that allow entrepreneurs—with alternative business models like the free textbooks supported by advertising, Bloomsbury Academic, Word Flat Knowledge, and various OER projects—to gain ground.

However, uncertainty regarding monetizing new free-content models (as seen in the newspaper industry), along with the perception of massive piracy in the music and film industries, has led to traditional publishers seeking ways to deliver textbook material through proprietary digital platforms or sell e-textbooks copyright-protected by digital rights management. What remains unclear is whether the market logic of the major publishers is ideologically incompatible with the open, unregulated commons, and which balance of economic, political, and social factors represents the greatest barrier to such innovation.

*XI - Next Steps*

Thus far, our research has been focused on the economic and political/regulatory factors affecting market innovation trends in the Higher Education and K-12 textbook markets, parallel to the acknowledgement of the existence of possible dynamics regarding other types of resources. We need to next look at the social factors/barriers affecting the producing actors, which we hope will give us a better sense of these markets' dynamics. To achieve this we propose a survey of teachers, similar to those undertaken by the Student PIRGs, that will look at issues of trust in traditional publishers and traditional EM formats, and other reasons for a lack of greater adoption of OER.

One particularly promising research opportunity is California's gubernatorial initiative to create open source digital textbooks for K-12 math and science. This could be a paradigmatic case in the effect of state adoption policies and public support moving the market toward less regulation and more open production. One step in this direction is structured interviews with teachers connected to the Community College Consortium for Open Educational Resources and the Community College Open Textbook Project in California. As the plan develops and various EM actors respond to the initiative, we can test how our Quadrants tool maps this evolving textbook market landscape and actors' strategies toward openness or closedness.

Lastly, our descriptive research exercise still needs to broaden. We plan to cover the "long tail" market of upper level Higher Education EM that involves significant production on the part of university presses. The alternative business models and classroom use of OER are more varied in this segment of the market, and the ideologies of more traditional publishers like some university presses can be less polar with respect

to open and closed models, as indicated on our current Quadrants mapping. Educational software is still an important segment that needs to be examined, particularly in light of bundling practices and new digital publishing models that deliver embedded textbook-like content through software platforms.

Another interesting exercise would be a deeper analysis of the commons-based cases observed, mainly among OER projects, in order to identify the characteristics that unit them around the openness side of the quadrant and which of those can be relevant for long-term sustainability of this approach, when comparing these efforts to well documented open-source efforts.

We also hope to expand the mapping exercise to other types of educational resources identified and improve the political economy mapping by adding lobbying groups, such as industry associations.

**ICP PROGRESS REPORT ANNEX 5**
**Innovation in Telecommunications**

*By Michael Steffen*

*I - Definition*

*"The word 'telecommunications,' a twentieth century amalgam of Greek and Latin roots, literally means the art of conveying information 'from a distance.' . . . Today, although precise definitions differ, 'telecommunications' is broadly defined as the transmission of information by means of electromagnetic signals: over copper wires, coaxial cable, fiber-optic strands, or the airwaves."[18]*

*II - Introduction*

Telecommunications technology touches every aspect of our lives. It affects the way we do business, the way we govern ourselves, the way we keep in touch with those we love, and the way we build the collective human experiences we call culture. Altogether, the telecom sector accounts for about fifteen percent of the U.S. economy.[19]

As outlined in Table 1 below, this paper explores one particularly dynamic area of change in the telecommunications industry: the ongoing broadband revolution in residential and mobile communication.[20] The nature of the telecommunications products and services that Americans use has changed dramatically over the last twenty years as a consequence of significant, sustained, and rapid innovation. This paper reviews these shifts, and then explores how the underlying innovation has come about, and in particular whether it has tended to follow proprietary or commons-based models. Have telecommunications innovators been driven to discovery by the promise of ownership over their discoveries, monetized through licensing revenue or by the exclusive sale of knowledge embedded products? Or have companies been driven to innovate in pursuit of a different set of rewards? If the latter, has the result been a commons in telecommunications technology available for harvest by others?

There are no simple answers. Different companies have adopted different models, and indeed a single company or academic institution may take different approaches depending on its strategic interests in particular negotiations. It is possible, however, to at least catalog the major approaches, and identify the forces that are shaping innovators' strategies.

---

[18] JONATHAN A NEUCHTERLEIN & PHILIP E. WEISER, DIGITAL CROSSROADS: AMERICAN TELECOMMUNICATIONS POLICY IN THE INTERNET AGE (2007), at 1-2.
[19] Nicholas Lemann, *The Chairman,* NEW YORKER, Oct. 7, 2002, at 48.
[20] This focus should not be taken to diminish the importance of enterprise telecommunications. In fact, company reports indicate that enterprise services are a larger share of revenue for the major telecommunications operators than residential services.

Part I of this paper provides a brief overview of the major ongoing changes in residential telecommunications driven by the rise of broadband. Part II connects these changes to areas of technological innovation, providing just enough background on network design to show what technological developments led to the rise of residential and mobile broadband. Part III reviews the basic value chain in telecommunications, describing the major players that have contributed to this innovation. With these background pieces in place, Part IV turns finally to the core questions of the paper, asking what incentives motivate the key players in telecommunications and how they manage their innovations. Finally, Part V concludes with proposals for further research.

### III - The Broadband Revolution

The major network owners that are the front line in the broadband revolution historically provided four distinct consumer-facing products: home telephony, mobile telephony, cable television, and internet access. In the residential market, these historical divisions are disappearing. Cable and telephone companies have each refashioned their networks to provide general-purpose high speed data transmission capacity. Using ever-growing and improving networks, both now compete to provide the dominant "triple play": telephony, television, and internet access. Municipalities and other new actors are building their own residential broadband networks, offering the same basic services.

Cell phone companies are also racing to become broadband providers. Cell phones have become much more than just phones, and data is rapidly overtaking voice as the dominant source of revenue in the industry. Mobile services offer lower bandwidth than residential service, and as a result, cellular networks will not be able to support robust wireless video for any substantial fraction of their users, and will not be able to support the same kind of "triple play" as residential broadband. But what mobile networks lack in speed, they make up for in ubiquity. Many analysts see the rise of mobile broadband as the most important and dynamic area in telecommunications in the short and medium term.

As Internet speeds and penetration increase—on both wired and wireless platforms—a new group of actors has also become increasingly important: so-called "over the-top" providers of communications services. Over-the-top providers are companies that compete with traditional telecommunications products and services over the public Internet—from the perspective of the traditional operators, these companies provide services "over the top" of basic consumer telecommunications, rather than as a component of the consumer package. Internet telephony companies like Skype and Vonage are the classic examples of this type of service, to which we also add makers of other innovative products and services used primarily for communication—things like email, online gaming, and virtual worlds.

The figures above and below demonstrate the revolutionary transformations ongoing in telecommunications based on the public operating data of the major U.S. carriers. Figures 1.1 and 1.2 show the growing role of cable companies in voice service and—more

recently—of telephone companies in video. As of the first quarter in 2009, Comcast announced that is the now the United States' third largest phone company, passing regional giant Qwest. Meanwhile, telephone companies have seen a rapid decline in the number of residential access lines they serve—more the result of losses to wireless subscribers who are "cutting the cord" than of losses to cable—but nevertheless a marked contrast with the rapid subscriber growth of the new entrants. On the video side, the rise of Verizon and AT&T as television providers is more recent and therefore less far along than the entry of cable into voice service. Thus Figure 1.2 shows only the last five quarters of video subscriber data as compared to the four years of changes in voice depicted in Figure 1.1. In this short time Verizon has not yet quite taken over the number five spot from Cablevision, but as in the voice market, the trendline is striking. Both Verizon and AT&T are quickly adding video subscribers while the largest cable companies have all been slowly shrinking or holding steady.

Figure 2.1 illustrates the effect that the convergence in service offerings, along with the growth in demand for high speed Internet, is having on companies' revenues. Just five years ago, video service accounted for four fifths of the subscription revenue received by Comcast, the largest U.S. cable provider. By 2008, the share was down to two thirds. The large phone companies do not break out their revenue in a way that makes a similar comparison possible, but based on the rapid decline in voice subscribers combined with steady growth in video and voice subscribers, we can surmise that they are seeing a similar diminution in the share of their residential subscription revenue realized from their legacy business.

Meanwhile, the mobile sector is also changing rapidly. Figure 2.2 illustrates the rise of mobile broadband. The share of revenue from data services realized by AT&T and Verizon (the two largest U.S. mobile providers) has grown from just 5% to over 25% in the last 4 years. This figure is somewhat overstated because cell phone companies count text messaging fees as data revenue—but even excluding these lucrative charges, analysts agree that the growth in the data side of the mobile business has been large and rapid.

Although all the above statistics are from U.S. companies, telecommunications providers worldwide are experiencing similar, fundamental shifts in their businesses.

### IV - Focus Areas of Innovation

For our purposes, the study of innovation in telecommunications is the study of the transformations described above. Technically, the various providers of new broadband services all offer some variation on the same very general network design. Fiber optic lines—by far the dominant modern telecommunications technology—form the high bandwidth core of any network. These glass cables can carry a quantity of information that is virtually limitless for all practical purposes. Backbone providers specialize in just this highest bandwidth segment of the network, in long runs between cities or underneath the sea. Other providers specialize in getting data from the backbone to end users, and some providers do both. Residential networks come in several varieties. In the case of a fiber-to-the-home (FTTH) network, fiber optic lines run all the way to the home. In fiber-

to-the-node (FTTN) networks, the fiber cable is stopped at a cabinet that serves a neighborhood, and data is carried from there to each individual home over legacy wires, typically twisted-pair copper telephone wires. Cable broadband networks are built on a similar design, with a few significant differences: the legacy infrastructure is coaxial cable, which is a higher bandwidth medium, but which is shared among the served houses. (In a telco-built fiber to the node network, each house has its own copper wire to the local fiber node.) In addition, each fiber node in a cable network generally serves on the order of 500-2000 homes, whereas each node in a telco FTTN network may contain a few hundred homes. Whatever the technology used to reach the home in a residential broadband network, the last leg within the home is often wireless, at least for the Internet portion of the broadband service. Cheap and widely available WiFi routers operate at low-power on open frequencies to provide this capability.

Although we often think of them as a fundamentally different technology, commercial cellular networks are not all that different from residential broadband networks: they are also just wired networks with a wireless last leg. Like residential networks, cellular networks are built with fiber at the core. This fiber extends all the way to many cell towers. The remaining towers are connected by legacy copper and coax links. Sitting at the end of these wired links, each cell tower is the equivalent of a WiFi base station, but with coverage up to at least ten miles depending on the location and network design. No doubt, digital cellular technology differs in important ways from home WiFi technology: it is optimized for a combination of voice and data rather than pure data, it includes complex systems to support communication with fast-moving devices (e.g. a cellular handset being used in a car), it is designed to reuse radio frequencies more efficiently, and it is engineered to allow the wireless link to be seamlessly "handed off" as customers move between one cell and the next. The bigger differences are regulatory rather than technical, however: cell towers are able to cover a much greater geographic range then a WiFi router because they are operated at much high power. High power operation is possible because the towers transmit and receive data on frequencies where the operator has purchased an exclusive license to operate from the federal government.

Table 2 provides an overview and comparison of the basic fixed and mobile network designs. As the table implies, the three broad areas of innovation necessary for the deployment of residential and mobile broadband have been:

- the development of fiber optic communications technology;
- the development of new network standards to coax greater speeds and two way capacity from legacy cable and telephony systems; and
- the development of new high speed wireless communication systems for both high-power licensed and low-power unlicensed frequency bands.
- At the same time, a fourth area of innovation has both fueled and been fed by these other innovations, as discussed in the introduction. Namely:
- the development of new "over the top" communications systems offered by independent companies over the public Internet.

The remainder of this overview focuses on these four areas of technological change. Although our focus is on the residential sector, the same basic areas of innovations are driving enterprise services.

## V - Innovation Flows in Telecommunications

Figure 3 illustrates a highly simplified value chain for residential broadband providers. Component manufacturers provide the basic optical and electrical building blocks for telecommunications systems—things like lasers and chipsets. Equipment and subsystem manufacturers assemble these items into complete network components—things like cell tower radios and switching systems. Finally network operators build and manage complete networks, selling services to consumers and businesses. In addition, over-the-top service providers sell further products and services that operate over the Internet and supplement or substitute for those services offered by the network provider itself. Table 3 list examples of major actors in each category along with their 2008 revenues from telecommunications-related divisions.

The first three major areas of innovation described at the end of the previous Part—each in different segments of the physical network—emerge from the complex interaction between system operators and their upstream suppliers. These relationships are dynamic and situation dependent. Innovation is neither simply manufacturer-driven nor operator-driven. Rather, operators have a set of market imperatives and competitive pressures that lead them to seek specific capabilities from manufacturers. These needs may be communicated in informal interactions, in formalized requests for proposals, or collectively through various industry associations. At the same time, equipment manufacturers constantly strive to develop new products that anticipate coming needs or give providers new capabilities. To a certain extent, network operators also do their own R&D, in part through collaborative consortia. The industry advances through the interaction of this push and pull.

The relationship between vendors and operators is also heavily shaped by standards processes. Operators want assurance that they will be able to buy interoperable equipment for different parts of their network from different vendors, and vendors want the large markets and economies of scale that come from building to broadly accepted standards. For obvious reasons, different pieces of network technology have to interoperate to a greater degree than do different components in most other technology-intensive industries. Thus, all parties have significant incentives to support standardization. Once a technical standard is adopted, it imposes a profound, durable effect on the industry, determining the specifications that vendors build to, and the capabilities that system operators offer to end-users. A number of different organizations lead standards efforts, each with a different membership and focus that shapes its work. Table 4 lists examples of major industry associations, research consortia, and dedicated standards bodies, along with basic membership information and standards activity.

Innovation in over-the-top Internet-based services occurs somewhat differently than for in-network technology. In particular, innovation on the Internet can be driven more by

freestanding actors, because inventions are embodied in software code running on general purpose machines rather than in integrated, special purpose systems. Often, the same company engineers a piece of software and uses that software to provider consumer-facing services (e.g. Skype). For this reason, Internet-based providers to some extent compete with both network operators and equipment manufacturers. There is creative friction in this competition, but also the potential for mischief on the part of network operators (who, as Internet access providers, are providing the platform for their own competitors). This tension is the source of high profile policy debates over mandatory unbundling of broadband services and "net neutrality" regulations.

Finally, as for all highly innovative industries, public sector research contributes substantially to telecommunications R&D. Military and university research constantly feeds the innovation pipeline. Table 5 lists examples of significant technologies that have emerged in part from the public sector. In a 1993 MERIT/SESSI survey of large firms in the EU, 70% (17 of 24) respondents reported that publicly funded research in electrical engineering was extremely important or very important to their unit's technological base.[21] This figure was somewhat lower than for comparable public sector inputs in other industries (for example 85% of pharmaceutical industry respondents indicated that public sector biomedical research was extremely or very important, and 78% of computer industry respondents indicated that public sector electrical engineering research was extremely or very important).[22] Nevertheless, the public sector contribution to telecommunications is indisputably quite large in absolute terms.

### VI - The Economics of Intellectual Property in Telecommunications

We can now turn back to the questions posed at the outset. It is worth pausing briefly to present the issues in a slightly more systematic fashion. At the highest level, we are interested in two closely related questions: (1) Analyzing innovations as *outputs*, are telecommunications companies motivated to innovate because of proprietary control that they can exercise over these innovations, or by other benefits that do not depend on restricting access to the fruits of their ingenuity? And (2) Analyzing innovations as *inputs*, is access to new discoveries difficult to come by, or are new discoveries readily available to those who would seek to utilize or build on them? In each case, the former possibility reflects a proprietary innovation environment, the alternative is commons-based.

In general, there are three basic ways in which a company taking a proprietary approach to its innovations may limit access in the pursuit of profit (or, from the perspective of a downstream innovator, there are three basic ways in which the use of preexisting innovations may be limited): a company may restrict *who* may use its innovations, it may restrict *how* the innovation may be used, or it may charge *fees* for access to the innovation. We label these dimensions as "openness," "regulation," and "cost." Closedness and high cost characterize proprietary models, whereas openness and low cost

---

[21] ARUNDEL ET AL., INNOVATION STRATEGIES OF EUROPE'S LARGEST INDUSTRIAL FIRMS, at Table C-12 (1995).
[22] *Id.*

characterize commons-based models. The regulatory dimension is more complex, because regulation of the use of innovations may be used to extract value in proprietary models, for example where a patent owner restricts licenses by use in order to protect certain markets for its product. But regulation may also be used in commons-based models to sustain the commons itself, in the way that traffic rules maintain the utility of the roads. Such is the well-known approach of open source licenses like the GPL.

In telecommunications literature, openness and cost are the major foci of concern. Regulation is less widely discussed, presumably because innovations, where available, are not restricted in their use, or at least not in ways that inhibit development or downstream innovation. Following the existing literature, the analysis below also focuses on the dimensions of openness and cost. Is access to innovation in telecommunications restricted? And is it expensive?

The answers to these questions differ somewhat between in-network technologies and over the top technologies, so the next two sections address each in turn.

## VI.1- In-Network Technologies

Telecommunications equipment manufacturers patent heavily.[23] Telecommunications system operators also patent, but apparently somewhat less so. Table 6 shows the total number of 2007 U.S. patents granted to leading system operators and equipment companies compared to biotech/pharmaceuticals companies and computer systems and software companies. These data must be read with due caution because some companies have units that fall into more than one category and because many factors affect the number of patent grants that have little to do with the extent of actual legal protection acquired—but the counts at least provide a rough indicator of the degree of patenting activity. One reason that system operators may patent less than equipment manufacturers is that operators achieve their margins by being in extremely capital intensive industries rather than through intellectual property. They exist in monopoly or oligopoly environments thanks to the economics of fixed costs, not because of government-granted rights to restrict use of their inventions.

A more systematic look at patenting activity in telecommunications is provided by Cohen et al.'s report on the comprehensive 1994 Carnegie Mellon Survey on Industrial R&D in the United States. Directors of research labs for telecommunications equipment manufacturers that participated in the survey reported that they filed patents on 60% of all product innovations, well above the cross-industry average (49%).[24] Using data from the 1993 MERIT/SESSI survey of large European firms mentioned earlier, Arundel and Kabla reached a similar result. Weighted by total sales volume, communications

---

[23] Software copyrights and rights in semiconductor designs, a *sui generis* form of IP, may also be important in certain instances, but patents are the most contested forms of legal protection in the industry and the focus of the most active legal and policy debate.

[24] Wesley M. Cohen, Richard R. Nelson & John P. Walsh, *Protecting Their Intellectual Assets: Appropriability Condditions and Why U.S. Manufacturing Firms Patent (Or Not)*, NBER Working Paper 7552, at Table A1, http://www.nber.org/papers/w7552.

equipment manufacturers reported that they patented on average 47% of product innovations, as compared to a cross industry average of 36%.[25] Although the specific percentages differ between the two surveys, the qualitative finding of above-average patenting is consistent. Again, the story may be somewhat different for system operators, but unfortunately Cohen et al. do not report data for telecommunications service providers, and Arundel and Kabla report data only aggregated with providers of physical transportation providers.

Notwithstanding high levels of patenting, makers of telecommunications equipment did not see patents as the most important means of protecting or monetizing innovations in either the Carnegie Mellon or MERIT/SESSI surveys. In fact, respondents to the Carnegie Mellon survey rated patents as the least effective among the specific surveyed means of appropriating value from new innovations, scoring behind lead time, secrecy, complementary sales, and complementary manufacturing. Patents scored low across all the industries surveyed, but telecommunications stood out even in the context of this general finding: the importance of patents was rated as far lower in telecommunications than in the cross-industry mean. Table 7.1 reproduces these data with comparisons to selected other industries.

As in other industries where widespread patenting activity accompanies a low perception of patent value, the primary cause is the prevalence of overlapping patent claims. Multiple patents, generally owned by different companies, are required to assemble a finished product. For example, the 3G Patent Platform Partnership estimates that over 100 companies own patents that are essential to implement 3G mobile telephony standards.[26] In such an environment, companies must patent widely at a minimum to protect their own freedom to operate: a strong patent portfolio allows a company to deter infringement with the threat of countersuits, but a company without a defensive portfolio is at the mercy of would-be litigants.

Because companies hold a mutual litigation threat, cross-licenses are common. Fourteen or fifteen (74-79%) of the nineteen communications equipment industry respondents in the Carnegie Mellon survey reported that they used patents in negotiations, to prevent infringement suits by other companies, and to block other firms from patenting related inventions. One respondent interviewed by the study's authors described the situation this way: "Mostly your patents are used in horse trading. . . . In our industry things all build on each other. We all overlap on each other's patents. Eventually we come to some agreement: 'You can use ours and we can use yours.'"[27] Table 7.2 shows the full survey results with comparisons to other industries. Arundel et al. report qualitatively similar findings from the MERIT/SESSI survey.[28]

---

[25] Anthony Arundel & Isabelle Kabla, *What Percentage of Innovations Are Patented? Empirical Estimates for European Firms*. 27 Research Policy 127, 133 (1998).

[26] Ky P Ewing, Jr, *EC and DoJ approval of the 3G Patent Platform*, GLOBAL COMPETITION REVIEW 12, Feb. 2003, *available at* http://www.3glicensing.com/articles/03%20-%203G%20(p12-14)%20f.pdf

[27] Cohen et al., at 19.

[28] ARUNDEL ET AL., at Table C-19a.

Standards processes also heavily influence the handling of IP, increasing the pressure to license broadly. As described in the previous section, standardization activity is central in telecommunications. Standards bodies generally require that participating companies disclose all intellectual property they own that is necessary to implement any new standard, and that the companies commit to license all such IP on "reasonable and non-discriminatory" (RAND) terms. (European standards bodies often add an additional obligation of "fairness"—making the acronym FRAND—but it is unclear whether this change actually adds any legal content.) Unlike patents that are only made available in the context of cross-licensing negotiations, patents that are subject to RAND obligations are available to companies that wish to enter the market even if they do not bring their own IP portfolio. Especially where only a small number of companies own the core IP, cross-licensing regimes can effectively perpetuate oligopolistic market structures. RAND commitments preclude such barriers to competitive entry. Accordingly, RAND licensing commitments provide some assurance to system operators that they will not get locked into just one or two suppliers. Meanwhile, equipment manufacturers are willing to submit to these requirements because of the huge scale advantages of having their IP included in a widely adopted standard. Indeed, if a company's IP is left out of an industry standard process that subsequently achieves dominance, its innovation is likely to fade to irrelevance.

The principle criticism of RAND agreements is that they are often vague and therefore difficult to enforce. As two lawyers with experience in licensing litigation put the issue:

> Standards bodies which make use of FRAND declarations—ie a promise that the licensor will make specified technology available on fair, reasonable, and non-discriminatory terms—rarely, if ever, give any guidance as to what those terms mean. . . . Nor will most standards bodies intervene in bilateral disputes between members (or between members and non-members) to set a FRAND royalty, or even to give any guidance on the meaning of the commitment, not the least because most standards bodies are little more than the sum of their members, with inevitably disparate commercial views. As discussed above, there is a significant risk that [FRAND or RAND] can mean all things to all men. In consequence, the obligation risks becoming toothless.[29]

Ambiguity in the meaning of RAND commitments leads to high transaction costs and lower transparency in the handling of IP, since each license must be individually negotiated (albeit often after a product is on the market, not before). Demonstrating the effect of these problems on downstream innovation, some industry players blame the failures of RAND licensing in part for limiting the spread of wireless technology to gaming consoles, smart energy meters, parking meters, and other new devices. As one executive at Intel stated the problem in 2008, "We haven't seen a broad proliferation of

---

[29] Pat Treacy & Sophie Lawrence, *FRANDly Fire: Are Industry Standards Doing More Harm Than Good?*, J. INTEL. PROP. L. & PRACTICE , Dec. 5, 2007, at 22.

cellular technology in anything other than handsets because the model is closely held and restrictive."[30]

Responding to the limitations of RAND agreements, some industry players are increasingly seeking to push IP policies towards greater openness, either by seeking specificity in the commitments made by standards contributors or by forming patent pools with standard in- and out-licenses. For example, several of the major players in the high speed WiMAX standard have formed a patent pool in an effort to "stimulat[e] a larger WiMAX industry that supports innovation through broader choice and lower equipment and service costs."[31] Similarly, many players in the various 3G mobile standards have banded together to form the 3G Patent Platform, a system for standardizing licensing terms designed to make licensing of 3G related patents simpler and more predictable.

The fact that telecommunications patents tend to be licensed broadly—whether through cross licenses, RAND commitments, patent pools, or otherwise—does not mean that these innovations are free in the economic sense. Nine of the nineteen communications equipment industry respondents in the Carnegie Mellon survey (47%) saw licensing revenue as a motivation to patent. Across all industries, only 28% of firms cited licensing revenue as a motivation for patenting. In other words, licensing revenue in telecommunications is substantially less important than the defensive motivations described above, but it cannot be ignored. In addition, many industry observers report that telecommunications companies have increased their emphasis on licensing revenue in the fifteen years since the Carnegie Mellon survey.[32]

The pursuit of licensing revenue varies greatly among component and equipment equipment manufacturers depending on the balance of their own IP and the IP of other players that goes into the products they produce (or whether they produce products at all). Qualcomm is a well known example of a company that assembled a sufficiently strong and free-standing patent portfolio to demand significant royalties in licenses for early digital wireless standards, and constructed a business strategy with a heavy emphasis on licensing. In 2008, Qualcomm reported $11.1 billion in revenue, of which $4.0 billion (36%) derived from licensing and royalty fees. An even more dramatic example is InterDigital communications, also a significant patent holder in advanced digital wireless technologies. InterDigital's 2008 Annual Report listed $229 million in total revenue, of which $217 million (95%) came from patent royalties. In contrast, licensing revenue does not merit its own line in the annual reports of companies like Cisco and Alcatel-Lucent, and these reports mention IP-litigation risk in the context of concerns that the company may be sued for infringement, rather than the possibility that a lucrative patent will be invalidated. Notwithstanding these generalizations, the interests of companies shift in different circumstances. For example, Alcatel-Lucent became infamous in 2006 and 2007

---

[30] Marguerite Reardon, *WiMax Patent Alliance Announced*, CNET NEWS BLOG, June 9, 2008, http://news.cnet.com/8301-10784_3-9963352-7.html.
[31] http://www.openpatentalliance.com/.
[32] *See, e.g.*, KEVIN G. RIVETTE & DAVID KLINE, REMBRANDTS IN THE ATTIC: UNLOCKING THE HIDDEN VALUE OF PATENTS (1999).

for its aggressive enforcement of MP3 patents against Microsoft, winning a $1.5 billion jury verdict before having the judgement overturned by the court of appeals.

System operators seem to have a more uniform position towards IP than their upstream manufacturers. As they do for openness, operators generally appear to pull in the direction of lower prices. IP factors into their economic equation primarily as a cost rather than as a means to thwart competitive pressure, and therefore their usual goal appears to be to push prices down. The PacketCable specification developed by CableLabs, a consortium of cable system operators, demonstrates this dynamic. PacketCable specifies standards for IP-based voice services on cable networks. In conjunction with certifying the standard, CableLabs set up a royalty-free licensing pool for related IP. (As this example perhaps suggests, the interaction of standards-setting and pricing concerns creates complicated competition policy issues.[33])

Of course, there remain circumstances when system operators change their approach to IP, just like equipment manufacturing companies. For example, Verizon, Sprint, and AT&T each sued Internet telephony provider Vonage for patent infringement in 2006 and 2007, extracting combined settlements of $240 million.[34] Then, in early 2008, Verizon sued two cable companies, Cox and Charter, over the same eight voice-over-IP patents that it had successfully asserted against Vonage.[35] Charter appears to have been better armed for battle than Vonage, however, and in December it fought back against Verizon, suing for infringement of four video and data transmission patents of its own.[36]

Figure 4 summarizes the discussion thus far, charting the basic licensing models used for in-network technologies in telecommunications on the dimensions of openness and price. Patent pools and other forms of standardized agreements are more open than IP licensed subject to RAND commitments, which in turn are more open than cross licensing arrangements. All these approaches can vary broadly in terms of the attendant pricing strategy. Figure 5 charts illustrative examples of some of these different strategies, many of which have already been mentioned.

From a policy perspective, the variation in licensing models means that different companies—or the same companies in different circumstances—have varying degrees of interests that turn on having strong patents in telecommunications. Many innovative telecommunications companies license their IP widely with zero or near-zero royalties (willingly or unwillingly), thereby adopting an essentially non-proprietary model and contributing to a commons in telecommunications innovation. But there are also

---

[33] Indeed, the fear of running afoul of antitrust rules is in part the reason that standards bodies long opted for general RAND obligations rather than specific commitments on pricing and other competitive terms. *See, e.g.*, Peter Grindley, Mark Bezant & Daniel Ryan, *Patent Licensing and Standards Setting — IP Collides with Antitrust, in* LICENSING IN THE BOARDROOM 2008 (2008), *available at* http://www.iam-magazine.com/issues/Articles.aspx?g=68bb21ce-9dc8-488c-98be-c4986ef63921.

[34] *See* Dan Frommer, *Vonage (VG), AT&T (T) Finalize Patent Settlement*, BUSINESS INSIDER, Dec. 21, 2007 http://www.businessinsider.com/2007/12/vonage-vg-att-t-finalize-patent-settlement.

[35] *See* Victoria Slind-Flor, *Charter Communications, GM: Intellectual Property*, BLOOMBERG.COM, Jan 8, 2009, http://www.bloomberg.com/apps/news?pid=newsarchive&sid=aucTsLbHm5w8.

[36] *Id.*

significant counterexamples, companies that may ultimately be pleased to see their innovation used broadly, but that seek significant compensation in return. Because they extract revenue based on a proprietary approach to their discoveries, these companies have interests in strong IP rights in telecommunications innovation.

A final caveat to these results is that what makes a company money ex post and what motivates it to innovate ex ante are related but not identical questions. Suggestively, 95% of telecommunications respondents in the MERIT/SESSI survey (18 of 19) reported that the desire to create new products was "extremely" of "very" important in "influencing the types or magnitude of innovative activities undertaken," while only 4% (1 of 25) reported that the desire to "earn revenue from licensing products" was that important.[37] This result is likely distorted by the fact that the MERIT/SESSI survey focused on large firms and is now over 15 years old. But it is nevertheless striking.

## VI.2- Over the Top Services

Internet-based communications services are relatively new, at least on the timescale of major cross-sectoral studies of innovation, and rarely studied as a distinct industry sector. Consequently, it is difficult to draw generalizations about the way intellectual property is used by the various companies that provide services in this category. Nevertheless, it is possible to at least note the areas of major public focus and controversy—namely, software patents and software licensing by software-as-service companies—and to make some anecdotal observations.

As the multiple lawsuits against Vonage (described above) demonstrate, patents can be very important for over the top service providers just as for more traditional providers of communications services. The Vonage example also suggests that some Internet-based providers—especially newer or smaller players—may not be aggressively patenting themselves, and therefore may be poorly prepared to defend themselves from patent suits. Larger players, like Google and Microsoft, are known to patent widely. It is less clear how these patent portfolios are being used. There are fewer high profile examples of these companies adopting aggressive licensing strategies as compared to certain vendors of in-network technologies, but quantitative data is unavailable.

Because the innovations of over the top service providers are frequently embodied purely in software running on general purpose computers (rather than physical equipment or embedded microprocessors, for example), there has been substantial controversy regarding whether or not they should be patentable at all.[38] In the United States, so-called "software patents" are generally permissible. In Europe, the situation is more complicated, and proposals to strengthen protection for software patents have met strong resistance.

The reason that the patent debate has taken on an added degree of importance in the software context is that the barriers to entry are otherwise so low. Becoming a

---

[37] ARUNDEL ET AL.

[38] *See, e.g.*, http://stopsoftwarepatents.org/.

manufacturer of telecommunications equipment requires substantial economic and social capital. In contrast, a handful of programmers with very few resources and pre-existing relationships can launch a new online service company. Witness the success of Facebook, for example. Facebook was started by four Harvard undergraduates in 2004, and built in under five years into the number one social networking platform on the Internet. Early in its life cycle, this sort of enterprise almost certainly does not have the resources to invest heavily in patenting. Opponents of software patents fear that if building a strong defensive patent portfolio becomes a prerequisite to release of innovative software, including innovative online communication services, the rapid innovation that has been the hallmark of the Internet to date will be dramatically slowed.

Significant attention has also focused the copyright and trade secret policies of online communications providers.  In general, the free software and open source movements have proven very successful in recruiting companies to contribute to open source products. But the open source model is unsettled as applied to software-as-service companies, which includes many Internet-based communications providers. Until recently, the requirement in "viral" open source licenses that users of open source software contribute improvements back to the commons was triggered by distribution of new versions. As a result, companies like Google and Facebook, whose custom software runs exclusively on their own servers and is never released to the public, were historically not bound by the obligations of open source licenses. But some members of the free software and open source communities began to feel that these companies are unjustly taking advantage of free and open source software (which they build upon) without contributing back. From the perspective of advocates, software as service companies have adopted closed, proprietary models of software development, albeit protected by secrecy rather than copyright or patent.

In 2007, the Free Software Foundation released a new variant of the Gnu Public License (GPL), called the AGPL, aimed at this concern. The Free Software Foundation "recommend[s] that developers consider using the GNU AGPL for any software which will commonly be run over a network." Any company that builds upon software licensed in accordance with the AGPL to provide online services is required to release the source code for its improvements, allowing others to adopt and build further upon those changes, but perhaps undercutting its own competitive advantage. Understandably, the desirability and viability of this new requirement have been subjects of widespread debate.

Because they are so young, the business models of Internet communications services are still undergoing rapid change.  As in the telecommunications sector as a whole, there are forces pushing both towards closedness and openness, towards proprietary models for the management and exploitation of innovation and towards commons-based models. It is still far from clear where a stable equilibrium exists.

*VII - Next Steps*

This paper has sought to provide a broad overview of innovation in telecommunications. But it is necessarily just a beginning.
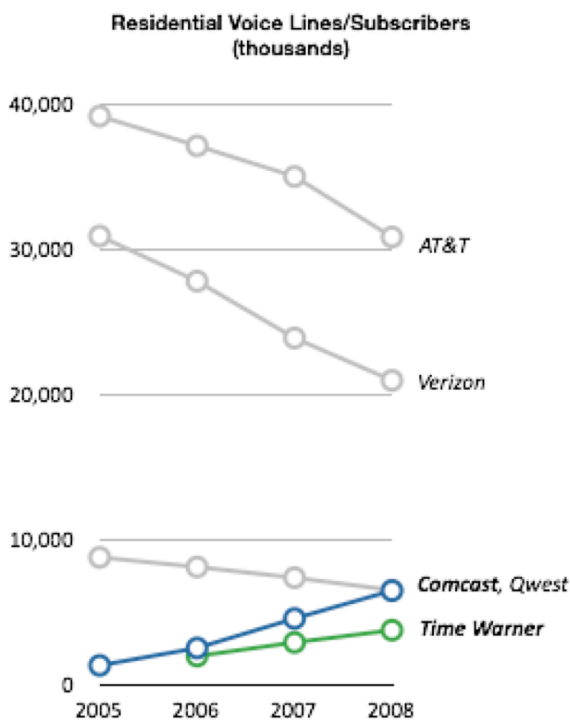
The most immediate next requirement is a series of telecommunications case studies to explore different examples of innovation in action. Ideally one would seek examples of representative innovations in each network segment and at each network layer: for example one might seek case studies in each of the fiber optic, legacy infrastructure (phone/cable), and wireless network segments, and at the physical, logical, and software/service layers.  Candidates might include erbium doped fiber amplification, a critical contribution to long-haul fiber networks; discrete multitone, a breakthrough that allowed efficient use of legacy telephone networks for DSL service; code division multiple access technology (CDMA), a transmission technology that revolutionized mobile data services; and Google Voice, an aggressive entry into telephone service by one of the largest Internet-based telecommunications companies.

Further research is also needed to add a political dimension to the analysis. What position have various telecommunications companies taken on the major intellectual property battles of the day—for example in Federal Circuit and Supreme Court cases governing patent standards or the availability of injunctions, or in legislative negotiations over patent reform? How have trade associations engaged in these debates? Are smaller companies that cannot afford (or choose not to invest in) individual political representation having their interests well looked after in centers of power? Placing an analysis of these questions in dialog with the economic research presented here and proposed for future case studies would allow a comparison of how companies perceive their self interest against what the economic data predicts.

Given the central role of telecommunications in the global economy and in the lives of humans worldwide, an understanding of innovation in telecommunications is critical to understanding the global dynamics of innovation generally. The technical, economic, and political dynamism of the sector means that there could be no better time for this work.
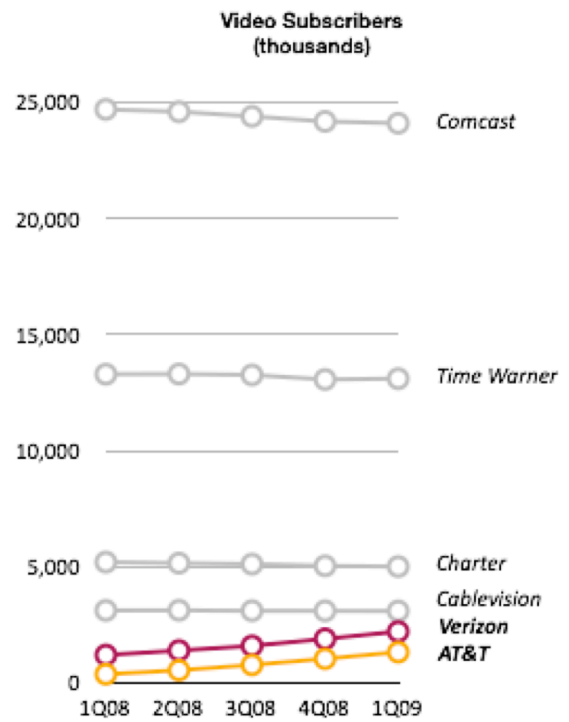
| Table 1: Field Definition | | |
|---|---|---|
| **Focus Industry Segments** | **Other Industry Segments** | **Excluded From Definition** |
| Wireline (Fixed Access Telephony) | Broadcast TV | Pure Content, including: |
| Cable | Satellite TV | Television and Radio Programmers |
| Commercial Wireless (Cell Phone) | Broadcast Radio | Online Content Platforms |
| Unlicensed Wireless Data (esp. 802.11) | Satellite Radio | (as distinguished from Communications Platforms) |
| Internet-Based Communications Platforms (e.g. Skype, email) | Other Wireless (e.g. public safety radios, maritime radios, cordless phones, etc.) | (e.g. Hulu, iTunes, Netflix) |



Figures 1.1 & 1.2: Voice and Video Subscribers

**Comcast Cable Segment Subscription Revenue**
**(percent)**



| | | | | | | |
|---|---|---|---|---|---|---|
| Internet | 15% | 19% | 22% | 24% | 25% | 25% |
| Voice | 5% | 4% | 4% | 4% | 7% | 9% |
| Video | 80% | 77% | 75% | 72% | 69% | 66% |

Source: Company reports

**AT&T & Verizon Wireless Service Revenue**
**(avg. percent, approx.)**



| | 2004 | 2005 | 2006 | 2007 | 2008 |
|---|---|---|---|---|---|
| Data | 5% | 8% | 14% | 20% | 26% |
| Voice | 95% | 92% | 86% | 80% | 74% |

Source: Company reports
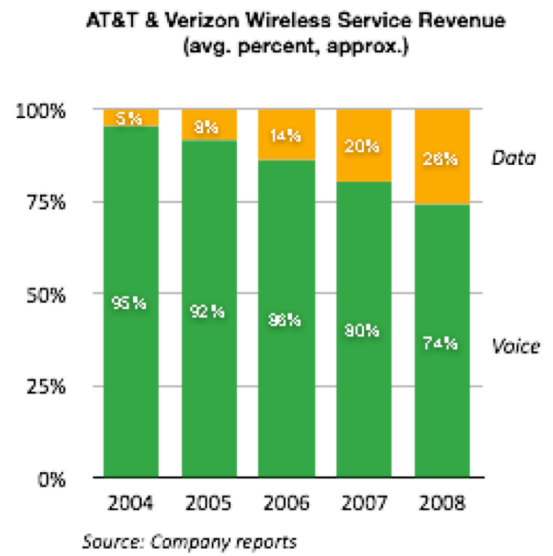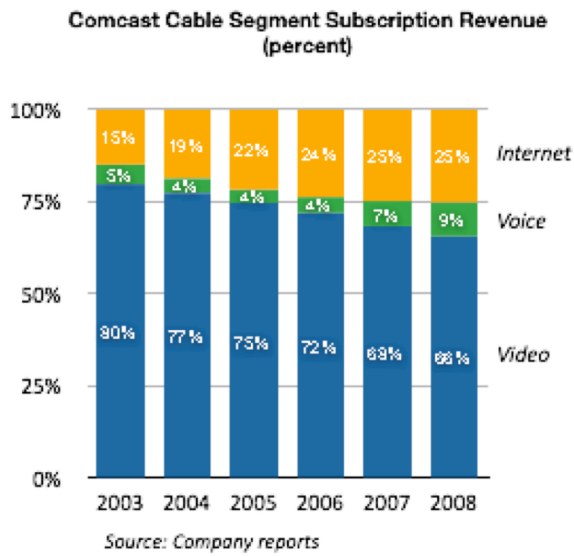
Figures 2.1 & 2.2: Cable and Wireless Revenue Sources

| | Fiber | Legacy | Wireless |
|---|---|---|---|
| **Telco FTTH (e.g. Verizon FiOS)** | to every home | in-home coax or telephone wiring (using MoCa, HPNA, or another standard) | in-home WiFi |
| **Telco FTTN (e.g. AT&T U-Verse)** | to a "node" (a node typically serves ~500-2000 households) | telephone wiring (using VDSL) | in-home WiFi |
| **Cable** | to a "node" (a node may serve a few hundred households) | coax cable wiring (using DOCSIS) | in-home WiFi |
| **Cellular** | in some cases, directly to towers, otherwise to multiple-tower aggregation points | "special access" lines to some towers from aggregation points | various digital cellular standards, depending on the network |

Table 2: Basic Network Components
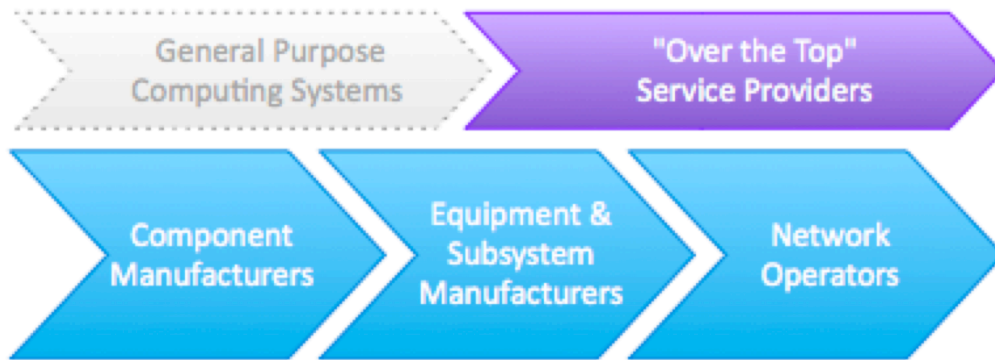


Figure 3: Basic Telecommunications Value Chain

## Table 3: Examples of Large Actors

| | major telecom products/services | 2008 revenue from specified activities/division |
|---|---|---|
| **Component Manufacturers** | | |
| Qualcomm | mobile chipsets | $ 7.5 B (QCT & QWI divs.) |
| Broadcomm | chipsets | $ 4.7 B (all divs) |
| Intel | wireless chipsets | $ 4.2 B (Mobility Group chipset revenue) |
| CommScope | cables, cabinets, antennas, electrical components | $ 3.8 B (all divs) |
| Texas Instruments | wireless chipsets | $ 3.4 B (Wireless div.) |
| Corning | optical fiber, cable, and components | $ 1.8 B (Telecommunications div.) |
| ADC Telecomms. | various components | $ 1.4 B (all divs) |
| **Equipment & Subsystem Manufacturers** | | |
| Nokia | mobile devices, networking systems | ~ $ 69 B (Devices and Services, Networks divs.) (50.4 B euros) |
| Cisco | routers, switches, networking systems | $ 38.0 B (all divs.) |
| Alcatel-Lucent | various networking systems | ~ $ 24 B (all divs) (17.0 B euros) |
| Motorola | mobile handsets, consumer premises equipment, networking systems | $ 22.2 B (Mobile Devices, Home and Networks Mobility divs.) |
| Huawei | various networking systems | ~ $ 18 B (all divs) (press accounts) |
| **Network Operators** | | |
| AT&T | U.S. fixed & mobile networks | $ 124.0 B (all divs.) |
| NTT | Japan fixed & mobile networks | ~ $ 109 B (all divs)  (10.7 T yen) |
| Verizon | U.S. fixed & mobile networks | $ 97.4 B (all divs.) |
| China Mobile | China and Asia mobile networks | ~ $ 60 B (all divs) (412 B yuan) |
| Sprint | U.S. primarily wireless network | $ 35.6 B (all divs.) |
| Comcast | U.S. fixed cable network | $ 32.4 B (Cable div.) |
| **Over the Top Service Providers** | | |

| Table 3: Examples of Large Actors | | |
|---|---|---|
| Vonage | internet telephony | $ 900 M |
| NewsCorp | MySpace social networking site | ~ $ 500-700 M (analyst estimates) |
| eBay | Skype internet telephony and chat | $ 551 M (Communications div.) |
| Yahoo | #1 webmail provider, chat | *communications revenue not separately reported* |
| Microsoft | #2 webmail provider, online gaming | *communications revenue not separately reported* |
| Twitter | short text messaging | *privately held (but revenues thought to be small)* |
| Linden Labs | SecondLife virtual world | *privately held* |

*Sources: Yahoo Finance,* http://finance.yahoo.com/; Company reports; Hitwise, *Top 20 Websites*, http://www.hitwise.com/datacenter/main/dashboard-10133.html (webmail rankings); Debra Aho Williamson, *Social Network Revenues Down: Here's Why,* EMARKETER.COM, (Dec. 23, 2008), http://www.emarketer.com/Article.aspx?R=1006825 (MySpace revenue estimate).

| | description/membership | examples of important standards/ areas of important standards activity |
|---|---|---|
| **Table 4: Examples of Significant Standards-Setting Bodies** | | |
| **European Telecommunications Standards Institute (ETSI)** | ETSI is an officially recognized but independent organization responsible for standardization of information and communication technologies within Europe. Its standards can have quasi-legal force. Its membership includes network operators, manufacturers, and some government bodies. | GSM European mobile phone standard part of 3GPP group developing third and fourth generation mobile standards |
| **International Telecommunications Union (ITU)** | The ITU is a longstanding UN body. Membership in its standardization activities consists of UN States along with companies as "sector" or "associate" members. The ITU sometimes originates standards activity, but also often approves standards after they have been first adopted by another industry body in order to give them international credibility. | Passive Optical Networking (PON) Synchronous Digital Hierarchy (SDH) wavelength-division multiplexing (WDM) Digital Subscriber Line (DSL) |
| **Institute of Electrical and Electronics Engineers (IEEE)** | IEEE is an international non-profit professional organization with a large standards-setting arm. IEEE allows both individual and corporate memberships, with standards processes following different paths depending on which class of membership is voting. | Ethernet (802.3) WiFi (802.11) WiMAX (802.16) |
| **Bellcore/Telcordia** | Created after the 1984 breakup of AT&T, Bellcore provided joint R&D and standards-setting for its co-owners, the Regional Bell Operating Companies. The companies later sold the enterprise, which changed its name and now operates as independent private company. Telcordia still performs standards-like functions under the name of "generic requirements" specifications. | Synchronous Optical Networking (SONET) Digital Subscriber Line (DSL) |

| Table 4: Examples of Significant Standards-Setting Bodies | | |
|---|---|---|
| **CableLabs** | CableLabs is a non-profit research and development consortium that was founded in 1988 by cable television operating companies to help them match the systems innovation capabilities of Bellcore. Its members are all cable operators. | DOCSIS (Data Over Cable Service Interface Specification) PacketCable managed voice-over-IP standard |
| **Internet Engineering Task Force (IETF)** | The IETF is has no formal membership or corporate status, but consists of individual participants organized into working groups and discussion groups that focus primarily on core internet standards. | email (SMTP/POP/IMAP) domain name resolution (DNS) network configuration (DHCP) |

| Table 5: Examples of Important Public Sector Innovations | | |
|---|---|---|
| | public sector contributor | description |
| **erbium doped fiber amplifiers** | Southampton University | EDFAs are a technology for amplifying optical signals, critical to long-haul fiber optic cables (e.g. for undersea use). The first EDFA was demonstrated by David Payne at Southampton University in 1987. |
| **RSA cryptography** | MIT | A cryptographic algorithm first published in 1977 by Ron Rivest, Adi Shamir, and Leonard Adleman at MIT, RSA is used in numerous settings where secure communication is required. For example, RSA is used in the DOCSIS standard to ensure privacy on shared cable networks. |
| **code division multiple access** | U.S. military | CDMA allows multiple radio devices to efficiently and robustly share the same radio frequencies in the same physical location. Developed by the military during World War II to frustrate jamming, CDMA was aggressively developed and popularized for commercial mobile use in the 1990s by Qualcomm. |
| **discrete multitone** | Stanford University | DMT is a technology to allow high speed communication over legacy copper telephone lines of varying length and quality, incorporated into the DSL standard. It was developed by John Cioffi of Stanford University, who founded a startup around the technology in 1991 and then sold the business to Texas Instruments six years later. |
| **Internet Protocol** | DARPA, Stanford University | The basic idea of a very simple but universally interoperable networking protocol gave birth to the Internet and has revolutionized telecommunications. It was first formulated by Robert E. Kahn of DARPA and Vinton Cerf of Stanford in a famous 1973 paper. |

| Table 6: 2007 Patenting Activity by Telecommunications Companies, With Comparisons | |
|---|---|
| **Company/Organization, Country** | **2007 U.S. Patents** |
| **Telecom Equipment & Services** | |
| 1. Siemens, Germany | 1305 7 |
| 2. Nokia, Finland | 30 705 |
| 3. AT&T, U.S. | 696 6 |
| 4. Alcatel-Lucent, France | 60 631 |
| 5. Cisco Systems, U.S. | 284 2 |
| 6. Motorola, U.S. | 77  27 |
| 7. Qualcomm, U.S | 4 228 |
| 8. Telefonktiebolaet LM Ericsson, Sweden | |
| 9. Nortel Networks, Canada | |
| 10. NTT, Japan | |
| **Biotech & Pharmaceuticals** | |
| 1. Roche, Switzerland | 515 47 |
| 2. Johnson & Johnson, U.S. | 6 283 |
| 3. Genentech, U.S. | 226 20 |
| 4. Pfizer, U.S. | 0 |
| 5. GlaxoSmithKline, U.S. | |
| **Computer Systems & Software** | |
| 1. IBM, U.S. | 3149 |
| 2. Microsoft, U.S. | 1649 |
| 3. Hewlett-Packard, U.S. | 1466 |
| 4. Fujitsu, Japan | 1490 |
| 5. NEC Corp., Japan | 972 |

*Source:* Patrick Thomas & Anthony Breitzman, *Patent Prowess*, IEEE SPECTRUM ONLINE, Dec. 2008, http://www.spectrum.ieee.org/dec08/7023.

## Table 7: Importance of Patenting & Its Causes

**7.1** *Mean percentage of product innovations for which each mechanism was reported effective in protecting "the firm's competitive advantage from those innovations" during the prior three years...*

| | N | Secrecy | Patents | Other Legal | Lead Time | Complementary Sales/Svcs. | Complementary Mfg. |
|---|---|---|---|---|---|---|---|
| Communications Equipment | 34 | 47% | 26% | 20% | 66% | 42% | 41% |
| Drugs | 49 | 54% | 50% | 21% | 50% | 33% | 49% |
| Medical Equipment | 67 | 51% | 55% | 29% | 58% | 52% | 49% |
| Computers | 25 | 44% | 41% | 27% | 61% | 40% | 38% |
| Semiconductors and Related Equipment | 18 | 60% | 27% | 22% | 53% | 42% | 48% |

**7.2** *Percentage of respondents indicating each reason as motivating their most recent decision to apply for a product patent...*

| | N | To Measure Performance | For Licensing Revenue | For Use in Negotiation | To Prevent Suits | To Prevent Copying | To Block Related Patents | To Enhance Reputation |
|---|---|---|---|---|---|---|---|---|
| Communications Equipment | 19 | 11% | 47% | 79% | 74% | 84% | 79% | 63% |
| Drugs | 36 | 14% | 44% | 61% | 67% | 100% | 97% | 69% |
| Medical Equipment | 60 | 5% | 22% | 58% | 65% | 95% | 93% | 57% |
| Computers | 20 | 0% | 30% | 80% | 90% | 85% | 65% | 40% |
| Semiconductors and Related Equipment | 12 | 0% | 42% | 67% | 67% | 92% | 75% | 33% |
| ALL | 765 | 6% | 28% | 47% | 59% | 96% | 82% | 48% |
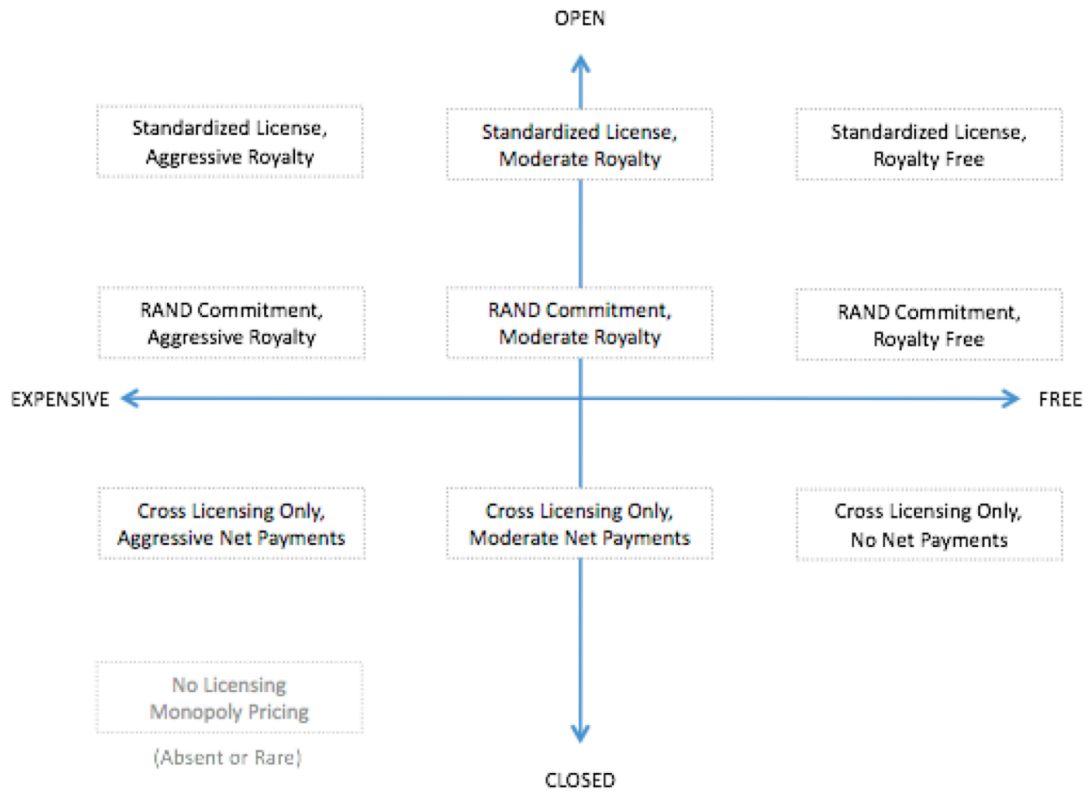
Figure 4: Licensing Paradigms in Telecommunications

OPEN

OpenPatentAlliance
WiMAX Patent Pool

PacketCable
Patent Pool

3G Patent Platform

WWiSE Group
802.11n Proposal

Qualcomm
Original CDMA Patents

Other GSM Contributors
GSM Patents

EXPENSIVE

Texas Instruments
DSL DMT Patents

FREE

Motorola
GSM Patents

CLOSED
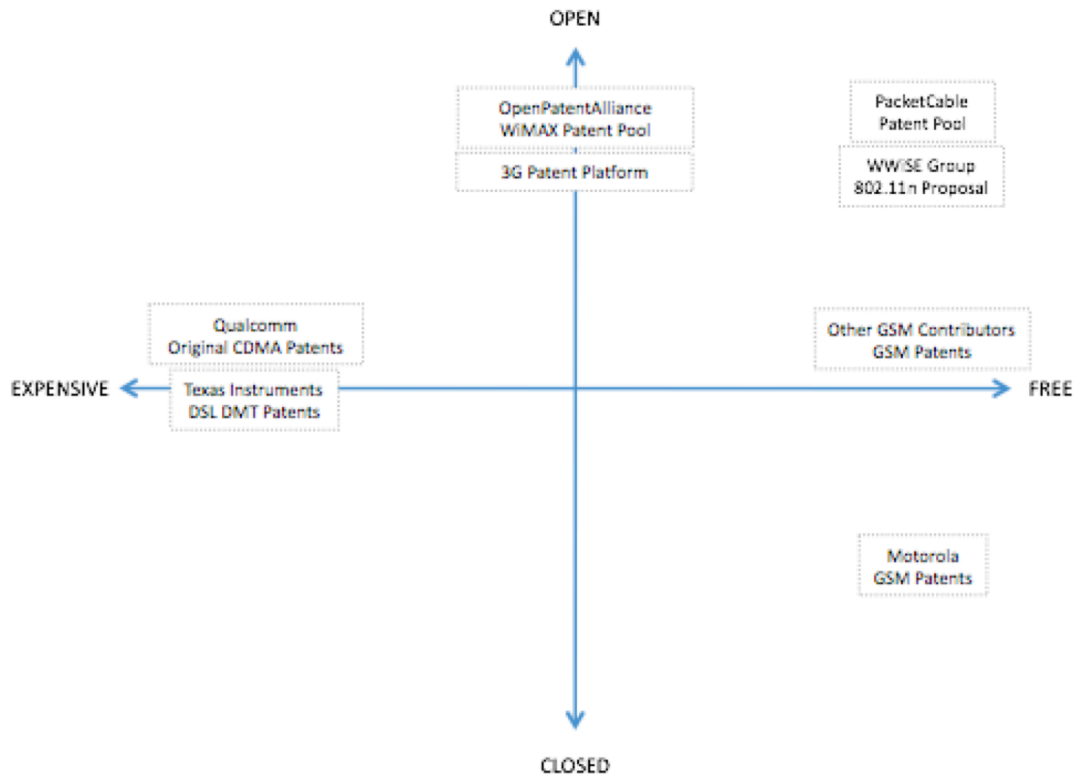
*Sources:* RIVETTE AND KLINE, REMBRANDTS IN THE ATTIC: UNLOCKING THE HIDDEN VALUE OF PATENTS (1999), at 125, 146 (Texas Instruments, DMT); DAVID MOCK, THE QUALCOMM EQUATION: HOW A FLEDGLING TELECOM COMPANY FORGED A NEW PATH TO BIG PROFITS AND MARKET (2005); http://www.openpatentalliance.com/; Department of Justice, Antitrust Devision, Business Review Letter on 3G Patent Platform, available at http://www.usdoj.gov/atr/public/busreview/200455.htm; Cable Labs, *PacketCable Royalty-Free IPR Pool Created*, Press Release, Oct. 30, 1999, *available at* http://www.cablelabs.com/news/newsletter/SPECS/specnewsoct/news.pgs/leadstory.html; WWise, *802.11n WWiSE Alliance Forms*, Press Release, Aug. 12, 2004, *available at* http://www.unstrung.com/document.asp?doc_id=57680; Rudi Bekkers, Bart Verspagen, Jan Smits, *Intellectual Property Rights and Standardization: The Case of GSM*, 26 TELECOM. POL. 171 (2002).

Figure 5: Licensing Examples in Telecommunications

# ICP Progress Report Annex 6

## A Case Study of Sage and the creation of the Public Domain

*Authors:*

*Carolina Rossini and John Wilbanks ([http://creativecommons.org/about/people#34](http://creativecommons.org/about/people#34))*

*Background for this paper:*

There are law review articles about the public domain aspect of life sciences, both as a government policy and as a precompetitive publication / patent defense. There are also articles about the use of collaborative, common platforms as vehicles for innovation.

In the life sciences, we find that the use of the public domain (if conformed to as a standard legal policy) can combine with convergent standardization on data systems and transform the data itself into a platform for innovation and collaboration. Thus, a default legal position by the government, or a defensive legal strategy by a large corporation, can have a "side effect" of enabling new forms of innovative R&D.

Our paper will analyze this effect with a specific focus on genomics.

*Skeleton for this paper:*

*The paper will cover a set of basic sections.*

1. The government created public domain in genomics.
   * Human Genome Project
     o Data access policy
     o Data access reality
     o Bermuda Rules etc.
     o Impact of Celera competition
   * SNPs and HapMap
     o Clickwrap license and intent
     o Corporate participation
     o end of clickwrap
   * ENCODE project
     o data access policy
2. The corporate created public domain in genomics
   * patent defense intent
     o enclosure by patent / startup / university tto
     o Merck Gene Index as "immune" response
   * alliance for cell signaling (AFCS)
     o what else?

3. technical platforms for genomics
    * Government centers
        o NCBI
        o EBI
        o Japan Genome
        o role of PD in global data integration
    * emergence of new PD tools
        o OBO
        o what else?
4. platforms for innovation
    * on NCBI
    * on Pubmed
        o Pubget
        o Hubmed
        o iHOP
        o Neurocommons / LOD
        o Ingenuity / Genstruct etc
    * on the genome
        o DAS
        o companies...
5. Why do some technical platforms make the transition to innovation platforms and not others?
    * information flow analysis
    * "resistance" analysis
    * Pubmed and genome v. AFCS and others
6. trying to recapture "lightning in a bottle" - the Merck experiment in disease biology (SAGE)
    * analyze decision in Merck's history with the gene index
    * analyze contracts
    * analyze technical platform

**Online Cooperation Research (OCR)**

# Contents

# Online Cooperation Research (OCR) Project

## I. Project Overview and Executive Summary

In the past twenty years, the Internet has given rise to networked communities that create public goods through the collaboration of thousands of individuals. Massive, distributed systems of social interaction, these communities challenge dominant theories of human organization and production. Wikipedia provides a dramatic, well-known example: since the creation of the site in 2001, people around the world have volunteered their time, effort, and knowledge to collaboratively write the world's largest, free multilingual encyclopedia. With little more than a skeletal social contract, the Wikipedia community members have created norms to govern collective behavior; divided their labor in complex ways; attracted a steady flow of new contributors; and produced thousands of articles of professional quality. The encyclopedia has become a near-ubiquitous resource, disseminating knowledge at a scale and speed that would not have been possible a short time ago. Why does this sort of distributed cooperative system work so well?

The rise of online communities like Wikipedia offers a unique opportunity to understand the foundations of cooperative systems and prosocial behavior. In addition, the variety and scale of cooperative activities online has grown to the point that systematic comparisons across communities are now possible. However, empirical understanding of online cooperation across domains and communities remains in its early stages.

Through the Online Cooperation Research project we aim to contribute to this emerging field of research through three distinct interventions: (1) the largest-scale observation survey of cooperative systems online to date; (2) a series of exploratory and analytical comparative studies that extend Benkler's theoretical framework of the design levers for cooperative systems and commons-based production; (3) the technical development of two open platforms for distributed and collaborative research: *Coopedia* and *ScriptGen.*

All three of these interventions have proceeded in parallel as distinct "phases" of our research agenda, illuminating each other and providing a broad empirical foundation upon which to base descriptive and analytical claims about online cooperative systems. Through Phases I and II, we aim to enrich our understanding whether or not differences in site governance, motivational framing, organizational structure, cultural norms, task design, and technological architecture determine the products of the cooperative systems in our study. We also aim to test a range of ancillary hypotheses about specific elements of the design lever framework. The platform development work complements these research activities by providing tools and techniques to facilitate subsequent studies in this area while also serving as a testing ground for many of our theories and hypotheses. The results of the OCR will contribute to the broader fields of commons-based production and human cooperative systems in a number of ways. Principally, the products of our research will greatly enhance the existing knowledge of this field,

providing a much expanded sense of the breadth and depth of the phenomenon than currently exists. These insights will help to inform both the subsequent design of cooperative systems as well as the application of effective policy and regulations to this dynamic space of innovation. In this way, the OCR complements the work of the ICP by expanding the scope of analysis beyond more traditionally recognized sectors of economic production into the relatively uncharted terrain of commons-based production online. Lastly, through the creation of free and open research platforms and new methodologies, we believe that the OCR will facilitate the future work of scholars and other enthusiasts in this field.

The rest of this report discusses each of these areas of activity – the large-scale observation survey, in-depth comparative studies, and research platform development – in greater detail, emphasizing the methods that we have deployed in the process as well as the research products that we expect to deliver. Given the fact that few resources or established methods of study existed that were appropriate for our project, much of our effort has gone towards producing a viable research plan and creating the tools with which to execute that plan. In this regard, our year has been a remarkable success as we have positioned ourselves to make a pioneering contribution to the field both in terms of the methods we have developed for this research as well as the preliminary results we have begun to gather. After explaining the different facets of the project in some detail, we offer a brief discussion of the time-line and results we expect to deliver. Each of the three phases is currently moving ahead at an appropriate pace for us to complete our data collection, analysis, and the launch of our research platforms within the coming year.

*II. Project Team*

**Host Institution:** The Berkman Center for the Internet and Society, Harvard University

**Principal Investigator:** Yochai Benkler

**Research Fellow and Project Manager:**  Aaron Shaw

**Project Coordinator:** Tim Hwang

**Research Assistants:**
> Manal Dia (Summer 2009)
> Yael Granot
> Scott Hartley (Summer 2009)
> Anna Y. Kim
> Roxanna Myhrum
> Ayelet Oz
> Dharmishta Rood

**Project Website:** http://cyber.law.harvard.edu/research/cooperation

### *III. Research Phase I: Large-Scale Observation Survey*

*Summary:*

The goals of our large-scale observation survey are to discover, validate, and analyze a generalizable sample of approximately one thousand cooperative systems online. We have designed original sampling and data collection methods to gather survey-style information about site governance, motivational framing, organizational structure; cultural norms, task design, and technological architecture. The results will allow us to compare variations across the communities and to test whether or not the factors listed above determine the products of the cooperative systems. In the process, we are developing replicable research methods and publicly accessible research platforms for future use by scholars and others.

The rest of this section presents (1) background to our study; the (2) sampling and (3) data collection methods of the large-scale observation study; (4) our plans for analyzing the resulting dataset; and (5) a description of our progress towards these goals.

*Researching Cooperative Systems Online:*

The topic of our research poses a number of challenges to traditional sampling and observation methods. These challenges derive in part from our unit of analysis – the cooperative system – as well as from the unique characteristics of the universe from which we draw our sample – the Internet. Building from examples and analysis developed in earlier studies, we begin from the following definition:

> ***Cooperative systems online*** *are entities with an online address (URL), including collaborative websites, projects, platforms, and communities, through which groups of people pursue activities that (a) deviate from the predictions of traditional theories of selfish rationality and action; (b) produce, distribute, or contribute to some shared (public) resource; or (c) otherwise engage in prosocial behavior or collective action.*

The definition is intentionally broad, consistent with the overarching theoretical framework and objectives of our project, as well as the range of cooperative activities that have emerged online. We also seek to synthesize the diversity of terminology used in existing research and popular literature on the topic. This diversity reflects the fact that the full extent of cooperative endeavors on the Internet is not only an *unknown* at this time, but may in fact be *unknowable*. The geometric growth of the World Wide Web coupled with the proliferation of the so-called the Dark Internet, the Deep Web, "darknets," private and semi-private intranets, and filtered, national networks has surpassed existing technical capacity to generate a reliable population from which to

draw a generalizable sample of URLs, domain names, or server IP addresses. Furthermore, even if it were possible for such a process to yield a representative sample of this kind, a prohibitive number of entities would need to be filtered before we could find a sufficient number of cases that met our criteria for cooperative human systems.

We have opted to draw a non-traditional sample from an alternative source: the current English language version of Wikipedia (abbreviated "en:WP" henceforth). In doing so, we harness the distinctive attributes of en:WP articles through a process we term "*Collective Intelligence Sampling*." We discuss this method as well as its limitations in detail below.

*Collective Intelligence Sampling*

One of the problems in generating a sample of Online Cooperative Systems is the lack of an authoritative set of definitions or lists from which to discover cases. The method of Collective Intelligence Sampling we have developed for this study overcomes these limitations by leveraging our existing knowledge of the field together with computational data-mining in order to discover a large sample of sites that meet the criteria for inclusion in our study. We define Collective Intelligence Sampling as involving the following four steps:

1. Identify a valid set of "seed" terms or cases that capture a range of concepts and examples of online cooperative hfuman systems.
2. Exhaustively crawl en:WP using an algorithm that scores the semantic relatedness of every article in the full en:WP set in relation to the terms of our "seed" set.
3. Use the results of the algorithm's crawl to discover the set of en:WP articles that meet a minimum threshold for relatedness.
4. Review these highly related results to identify and validate the cases of online cooperative human systems that meet the criteria of our definition (above).

This method builds on techniques originally developed for quantitative text analysis and qualitative content analysis. For the data-mining, we rely on an algorithm developed by David Milne and implemented in his Wikipedia Miner Toolkit to measure the semantic relatedness of any two articles in the English language Wikipedia.[39] At present, we have identified a "seed" set of approximately 200 terms and cases based on a literature review of relevant works in this field and are collaborating with Hal Roberts, a senior member of Berkman's technical staff, to adapt Milne's toolkit to our project (see Annex 2, "Collective Intelligence Sampling" abstract). We anticipate that we will complete the

---

39 Available for download and demonstration at: http://wikipedia-miner.sourceforge.net/ (Accessed May 13, 2009). The toolkit scores semantic relatedness based on measures of network centrality and proximity as well as the content of the "anchor text" that is highlighted in the hyperlinks within Wikipedia articles.

selection and validation of the highly related set of en:WP articles later this summer.[40] The result will be an exhaustive list of websites, online platforms, communities and projects that meet the standards for inclusion in our study.

The English language Wikipedia has several advantages and disadvantages as a source for information about online cooperative human systems. Wikipedia comprises one of the single largest repositories of information about the Internet and online phenomena. Unlike other repositories or listings of cooperative human systems online, en:WP is not constrained to a particular domain (such as Free Software projects) or a particular type of cooperative practice (such as social networking). A preliminary review of potential examples of online cooperative human systems that we conducted with our research assistants in the Fall of 2008 revealed over 1500 unique cases that met the criteria we described above.[41] The content of en:WP has also been collaboratively written and filtered by many thousands of contributors, and, as the product of such collective effort, achieves quantifiable advantages as a source for semantically structured information. In addition, en:WP has structural properties that make it possible to extract and analyze this information in useful ways: the site is a semantically organized, densely hyperlinked "small world" network in which the connections between pages signal meaningful relationships between the ideas they represent. These properties facilitate the application of automated data-mining techniques. Finally, complete archived copies of the encyclopedia are available under an open license, which facilitates subsequent replication and refinement of our work.

Sampling from the English language Wikipedia also imposes certain constraints on our study. The most important limitation – and the one that is most difficult to measure – is the extent to which the contents of en:WP are biased and do not include a representative sample of the objects of our study. For example, it is likely that non-English language cases that would meet our selection criteria have a lower chance of being included in the study because they are less likely to be included in en:WP. Furthermore, it may appear that we have "selected on the dependent variable" by drawing from a subset of the population where "successful" online cooperative systems are likely to be over-represented. We respond to these and concerns in several ways. The question of dependent and independent variables is fundamental, and illuminates the limitations of our study as we will not, on the basis of this survey, be in a position to infer the predictors of success or failure for online cooperative systems with any certainty. Our research questions reflect this issue insofar as we focus on comparative analysis across different facets of systems design, not across successful and unsuccessful systems. We also plan to include a subset of paired successes and failures among our smaller scale comparative studies to gain insights into these concerns.

---

40  The validation that the Wikipedia articles identified by the algorithm meet our criteria for online cooperative human systems will require human filtering. We will conduct this work on Amazon Mechanical Turk through the ScriptGen platform we are developing for subsequent stages of our observation survey.

41  We did not pursue this snowball sampling further because of the limitations it would have imposed on our subsequent analysis. By using Collective Intelligence Sampling, we will be in a position to generalize our findings in a more rigorous way – something that would not have been possible otherwise.

In response to the broader question about selection bias, we have undertaken a parallel effort to analyze the relationship of entities included en:WP and those that are not. While the examples of online phenomena included in Wikipedia may be *extensive*, we know that they are not *exhaustive*. Based on existing studies that have demonstrated a highly unequal "power law" distributions of attention and links in many domains of the Internet, we anticipate that Wikipedia will have predictable, measurable biases. For example, in the case of SourceForge.net, a site where Free and Open Source Software projects are hosted, there are over one hundred sixty thousand total projects, of which about four hundred also have an article about them in Wikipedia. Because SourceForge distributes data about all of the projects hosted on the site, we can compare the characteristics of the small set that have Wikipedia pages with the rest of the population. Our preliminary analysis indeed suggests a strong and predictable bias on Wikipedia for articles about SourcForge projects with high levels of activity and use. By better understanding the relationship between the two and by replicating this technique with other, similarly exhaustive data sets for related domains, we will be able to address the limitations of our sample. Aaron Shaw will pursue this analysis further together with Benjamin Mako Hill this summer (see Annex 2, "Wikipedia as a Window on the World?" abstract). This enhanced understanding of the constraints of en:WP will inform the extent to which we can generalize our results.

*Data Collection: Crowdsourced Observation Survey*

Once we have gathered and validated our sample of Cooperative Human Systems Online, we will conduct a survey to collect structured data about each case. To achieve this, we have designed an extensive survey instrument to assess attributes of each site relevant to our core research concerns, including questions about site governance, motivational framing, organizational structure; cultural norms, task design, and technological architecture (see Annex 1 for the latest version of the instrument). We have worked with our team of five Research Assistants during the past year to develop, pilot and revise the instrument. We are currently conducting a second round of pilot studies with the latest version, and aim to complete final revisions of the instrument later this summer.

Upon completion of the survey revisions and testing, we plan to administer the data collection through an online interface combining the Amazon Mechanical Turk (AMT, http://www.mturk.com) crowdsourcing platform and the ScriptGen tool, an open-source web-based survey application currently under development at the Berkman Center.[42] For each site in our sample, we distribute the questions in small, thematic batches that are randomly assigned to workers drawn from the AMT labor pool. We ensure that the

---

42  A brief explanation of AMT:  AMT, which is advertised as "artificial artificial intelligence," is an online labor market that facilitates the distribution of jobs to a large, decentralized community of independent workers. Through the AMT interface, job "requesters" can post a task and payment offer. Workers then peruse the task listings and choose which ones they want to accept. In general, the tasks tend to be small, repetitive tasks that are fairly easy and payments tend to be low (on the order of $0.10 per task). The extent of interaction between workers and requesters is minimal.

resulting data meets high standards and precision and accuracy through several means. First, we score workers performance against a small set of "gold standard" data validated by our own researchers ahead of time. This establishes a metric of each worker's performance and enables us to weight the rest of their responses in accordance with their score. Second, we also elicit overlapping data for every question and thereby generate a confidence interval around the responses as well as a means of filtering responses through the posterior application of Bayesian inferential models (Snow et. al. 2008).

This mode of distributed or "crowdsourced" observation survey research has only recently become possible and has never, to our knowledge, been attempted on a similar scale in the social sciences. Traditional survey and content analysis methods used a relatively small number of observers with similar training and measured success based on the extent to which these "coders" were able to agree reliably on accurate answers. More recently, researchers engaged in large-scale text analysis and natural language processing have expanded the scope of these approaches through the use of structured models and non-parametric inferential techniques.[43] We draw from both approaches by integrating the richness of human observation with the scalability of distributed, automated analysis. with From a methodological and logistical point of view, our approach promises several major improvements over earlier studies: By distributing the work via AMT, we reduce the possibility that correlation between observations might result from unreproducible inter-rater biases as was likely with traditional content analysis. Furthermore, by distributing our research across a large population of workers, we gain the ability to ask structured, qualitative questions about hundreds of websites – something which has still proven impossible by automated methods.

*Proposed Analytical Techniques:*

The data collection and analysis for the large-scale study will proceed on a rolling basis throughout 2009-2010. We will use the data-set that results from this study to test a wide range of research questions about the cooperative systems online. Examples of these questions include: (1) Do more participatory cultural institutions prevail in less hierarchical, non-corporate organizations? (2) Which technical and organizational features best predict whether a site's participants engage in complex, highly-skilled activities or not? (3) Among systems in which users make predominantly small, modular contributions (such as applying tags or ratings to content), do site leaders appeal to in-group solidarity more widely than other motivational rhetorics? Answering these questions will help us refine existing theories of networked communities and explain the relationships between the different features of cooperative systems design and the existing landscape of online cooperation.

Given the relative lack of comparative data in the field of cooperative systems online, our analysis will also include a more exploratory component. Drawing again on data-mining

---

43  See King, Gary and Daniel Hopkins. Forthcoming in *Journal of the American Political Science Association*, "A Method of Automated Non-Parametric Content Analysis for Social Science. Available at: http://gking.harvard.edu/files/words.pdf (Accessed June 1, 2009).

techniques prevalent in the field of natural language processing, will use unsupervised clustering algorithms and machine learning techniques to search for unanticipated relationships among the cases included in our study. This approach has methodological limitations insofar as it is often inappropriate to generalize on the basis of any findings that emerge through data-mining alone. Nevertheless, it is likely to help us reveal previously unexplored connections between different systems and their attributes so as to identify areas that would benefit from follow-up studies.

*Phase I Status and Time-line:*

Our efforts this academic year have focused on developing the appropriate techniques and tools for carrying out this research. In addition to writing and piloting the survey instrument, this has involved collaborating with computer scientists, web developers, and social scientists to implement our plans. As noted above, we expect to complete the collective intelligence sampling process in July and are also beginning pilot studies with our survey instrument this summer. Working with a software development team within the Berkman Center, we have customized, tested, and refined an web-based research tool (tentatively called ScriptGen) to administer the crowdsourced observation survey via Amazon Mechanical Turk. We are now conducting a series of experimental studies with ScriptGen to determine the optimal methods for survey administration on AMT.[44]

The ScriptGen development will continue during the summer months when we also aim to release it under an open source license to facilitate sharing and collaboration with other research groups and institutions (See section V of this report for more details about ScriptGen). We are in the process of documenting this research process and our methods more fully, with the aim of submitting the work for peer-reviewed publication during the 2009-2010 academic year. By the end of summer 2009, we aim to complete the testing stages of our work and begin data collection for our large-scale observation survey. Data collection for the survey will continue through 2009-2010, during which time we will also conduct preliminary analysis of the results as they become available. In addition, these results will inform our ongoing series of small-scale comparative and qualitative studies (see Phase II, below). Upon completing the data collection for the survey, we will conduct the analysis described above and produce a series of articles that address both specific research hypotheses as well as our general findings about the space of cooperative human systems online. Upon publication of these early reports, we will release the data-set and all of our research instruments under an open license to facilitate the evaluation, reproduction, and extension of our research.

---

44 For example, the first study we began running in early June measures AMT worker performance across different motivational frameworks. Drawing on earlier studies in experimental economics and social psychology, we test sixteen different treatments in a controlled experiment. Using actual questions from our observation survey instrument, we measure worker responses against highly-reliable gold-standard data collected be members of our research team. The results of this study will make an original contribution to the experimental literature on cooperation, motivation, and labor market psychology at the same time as it helps us refine our research methods for the larger-scale observation survey. See Annex 2, "Driving the Hive" abstract.

### IV. Research Phase II: In-Depth Comparative Studies

*Summary:*

We use Phase II of our study to pursue a series of more narrow research questions about cooperative human systems online through a collection of in-depth comparative studies. Within a rubric of comparative cooperative systems analysis, our Phase II studies encompass a broad range of topics and methodological approaches. Some studies have been conducted by our team of Research Assistants and serve primarily as exploratory complements to the Phase I research. Others are the product of focused collaborations with researchers in other academic departments and institutions.

In contrast with the large-scale observation survey, Phase II looks more closely at particular *cooperative problems*, that is to say, those problem(s) that a given system attempts to solve through cooperation. In analyzing cooperative problems, we aim to understand the structural features of salient cases and how those features affect collaboration on a site. Our comparative analysis thus builds from the survey work, examining similar aspects of cases at a much higher resolution. Finally, by linking studies across domains, we also consider how and why different environments might or might not share common design elements, technologies, and attributes. In this way, Phase II both extends and informs our work in Phase I by providing a much richer and more precise sense of the mechanisms, dynamics, and experiences that comprise the space of online cooperation.

*Current Phase II Studies*

The Phase II studies generally fall into one of two categories: exploratory or analytical studies. Within these two categories, studies may also be *domain-specific* (e.g. the domain of political blogs or hospitality communities) or *cross-domain* examinations of cooperative systems design elements. This is true of both the work we and our Research Assistants have completed internally as well as of the work we have conducted with outside collaborators. The topics for the studies range widely and include political news and discussion communities, hospitality exchange networks, "virtual airlines," crowdsourcing platforms, and Wikipedia itself. Many of the studies perform double-duty insofar as they also promise generalizable insights which we use to test and refine the methods for Phase I of our project (see Section III above). We include a list of titles here to provide a general sense of the topics we have addressed. We have also attached brief abstracts for all of the studies in Annex 2 as well as a selection of the exploratory studies in Annexes 3-5.

*List of Phase II Studies:*

- Collaborative News Reporting and Citizen Journalism Sites
- Hospitality Websites: Design Levers for Cooperation
- Virtual Airlines
- Collaborative Fashion Sites
- Collaborative Entrepreneurship and Business Communities
- Bad Guys: Illegitimate and Illegal Collaboration Sites
- The Daily Kos: A Case Study of a Large-Scale Collaborative Political Blog
- A Tale of Two Blogospheres: Discursive Practices and the Left and Right
- Driving the Hive: An Experiment in Motivation and Task Design in a Distributed Labor Market
- Wikipedia as a Window on the World? Notability and Bias in the World's Largest Free Encyclopedia
- Collective Intelligence Sampling: Mining Wikipedia With a Purpose

*Phase II Status and Timeline:*

During the 2008-2009 academic year, we have pursued much of our Phase II work with a team of research assistants. After devoting much of the fall to writing, testing, and refining the observation survey instrument, our team began Phase II by identifying a group relevant domains for investigation and then conducting a series of exploratory studies to gather descriptive information and direct subsequent analysis. The members of the team then submitted reports on the results of these exploratory studies and several of the researchers are now beginning follow-up analytical studies during the summer under the supervision of the project managers. We are also pursuing collaborations with other scholars with whom we aim to conduct similar small-scale or domain-specific comparison studies. The results of Phase II will be a series of article-length studies, peer-reviewed papers and reports which encompass a wide range of empirical evidence and analytical methods.

During the coming months, we will continue to develop a body of reports that compare cases within and across domains of cooperative systems online. Drawing on the exploratory studies that were completed this year as well as the more in-depth projects that we will pursue this summer, some of the reports will reflect an extended engagement with the design levers and cases included in the sample for our large-scale survey. We will also continue to develop independent studies with outside collaborators that complement our insights into the design levers, cooperative systems dynamics, and the different domains of networked cooperative action (see the abstracts in Annex 2 for examples). Finally, as we begin to collect data through the large-scale observation survey later in the year, we also plan to begin work on follow-up studies that will help us better understand the results and insights of Phase I. In this way, by the end of the next academic year, we aim to complete a body of 5-10 research papers of publishable quality that complement our work in the survey.

## V. Distributed and Collaborative Platform Development

*Summary:*

For the third and final component of the Online Cooperation Research project, we have developed distributed and collaborative research platforms. Our efforts in this area have focused on two discrete, but interrelated projects: Coopedia and ScriptGen. Each tool facilitates the observation and analysis of online cooperation by distributing tasks across a large group of contributors and/or research workers; however they do so in vastly different ways. Coopedia follows the model of large-scale volunteer communities (like Wikipedia) in order to elicit information sharing and meta-analysis about the structure, design, and participants in cooperative systems online. ScriptGen, by contrast, provides a flexible interface for researchers to assemble, distribute, and collect data through online surveys and experiments by a number of different means. Both tools extend and complement the other phases of our project by enabling us to conduct analysis on a much wider scale than would be possible otherwise. As a result of the fact that we will also release both platforms and, in the case of Coopedia, their contents under free/open-source licenses, they will also contribute to the overall enrichment of the commons. The rest of this section explains each platform in greater detail and provides an overview of their respective development timelines.

*Coopedia*

Coopedia facilitates the creation and sharing of information about online cooperative human systems. It is a web-based platform aimed at aggregating the contributions of two overlapping end-user groups: the participants and the observers of online cooperative systems. This platform offers a persistent format through which these end-users contribute to parallel pools of information – one a collection of reflections and notes written by cooperative systems participants; the other by observers or analysts of those same systems. The information pools are repositories of experiential and interpretive knowledge about online cooperative human systems, facilitating subsequent reflection, research, and innovation in this arena. As a result, Coopedia represents a practical extension of our research and builds on the project design we have used in both Phases I and II.

Based on an initial strategic plan we created for Coopedia during the fall 2008 semester, we have pushed back the active development of the tool until the late summer 2009. The reasons for this are two-fold: First, we determined that both the structure and initial body of the content would depend on the survey instrument we designed for Phase I as well as the results of the exploratory studies we have conducted in Phase II. Consequently, we prioritized the work on these phases. Secondly, we sought to initiate a collaboration with Professor Archon Fung of the Harvard Kennedy School who, under the auspices of the

Ash Institute for Democratic Governance and Innovation, is developing a similar platform to promote civic engagement and political transparency under the name of the Participedia project. At this time, both of these prior steps have been completed.

We are now ready to move forward more aggressively on the design and development of Coopedia. We have begun writing and refining a design specification together with several research collaborators and members of the Berkman Center's technical staff. Our goal is to finalize this specification early in fall 2009 and begin development on the platform shortly thereafter. As described above, the layout and architecture of the user-interface will reflect the structure of our survey instrument by incorporating the same categories of analysis and many of the same questions. Furthermore, at the time of its launch, Coopedia will be "seeded" with Wikipedia content for all of the cases included in our large-scale observation survey. We will supplement this content with observation data collected as part of our exploratory comparative studies. At this time, our goal is to have a working prototype of the site ready by December, 2009 with a full launch scheduled for April 2010.

*ScriptGen*

ScriptGen is an free and open-source online tool for administering surveys, experiments, and questionnaires. The software was originally created by Berkman staff members Jason Callina and Anita Patel for use by several projects at the center that involved the assignment of text-based scripts (e.g. a set of survey questions) to a pool of research subjects or assistants, each of whom has also been assigned to particular research objects (in our case, websites that are examples of Cooperative Systems Online). In this way, ScriptGen makes it possible to create and manage large-scale human observation tasks through a web-based interface. Distinct from existing survey or experiment software, it also facilitates randomized task assignment and integrates seamlessly with Amazon Mechanical Turk. The questions, stylesheet information, and data are all stored in a relational database and can then be exported in multiple formats for subsequent analysis, reproduction, or review.

ScriptGen has been under development since the Fall and can already handle a wide range of research tasks. At present, we are using it to conduct an experiment that is both a Phase II study on the motivations of workers in a crowdsourced labor market as well as a pilot test of some of our observation survey questions from Phase I on Amazon Mechanical Turk (see Annex 2, "Driving the Hive" abstract). Also, we are working to build additional features that will make it so that ScriptGen can handle the expanded set of questions and functions necessary for the full Phase I observation survey. Aaron Shaw and John Horton, a Harvard Kennedy School PhD student and collaborator on the study, recently presented this work at a meeting of Amazon Mechanical Turk researchers and entrepreneurs in San Francisco.

In conjunction with our preparations for the Phase I observation survey with ScriptGen, we also intend to pursue several other goals with this piece of the project in the coming

year. First, we are in the process of planning a series of follow-up pilot studies and experiments to test other subsets of the survey questions. By comparing the answers of Mechanical Turk workers against "gold standard" data that we will collect ourselves, these studies will help us ensure that the questions produce valid results. They will also provide ample opportunity to continue testing and refining the tools features. Along these same lines, we are also in the process of refining the feature-set and development time-line for the next few months of the project. This summer, Callina and Patel, together with an intern sponsored by the Google Summer of Code program, are adding critical features and streamlining the interface of the tool in order to facilitate its adoption by non-expert users. We are also preparing to release the tool under an open license later this summer and create a code repository so that other people will be able to download and contribute to its development. In the coming weeks, we expect to begin the process of attracting additional users and programmers to the platform by cultivating collaborative relationships with other research centers around the Harvard campus and beyond.

*VI. Conclusion*

In many respects, the OCR began with a series of seemingly intractable bootstrapping problems. The project has now moved past that stage to the point where we have the appropriate tools, methods, and objects of analysis necessary to achieve our objectives. During the last ten months, we have laid the groundwork for an unprecedented large-scale observation study in the form of Phase I; an expansion and deepening of the sort of in-depth and comparative reports we have begun with our Phase II work; and the launch of Coopedia and ScriptGen, two separate free and open web-based platforms for distributed research and information-sharing. We anticipate that these investments of our time and resources will pay substantial dividends in our second year.

# OCR Progress Report Annex 1

## Observation Survey Instrument

**Sections:**
   A. **Features and Content**
   B. **Legal**
   C. **Assessment and Tags**
   D. **Organization and Leadership**
   E. **Norms and Enforcement**
   F. **User Behavior**
   G. **Motivations**
   H. **Tasks**

## A. Features and Content

1. Does the site use any of the following social features? (Please indicate yes or no for each feature)
   a. user identities or usernames
   b. user profiles or pages
   c. news feeds (such as RSS or Atom feeds)
   d. avatars or other visual identities
   e. discussion forums
   f. collaborative writing or editing tools (like shared pages or a wiki)

2. Do users of the site have visible reputations? (For example, a user might have a high reputation score based on their past contributions) If you are unable to find an answer or if the question does not apply, leave the response blank.
   a. Yes
   b. No

3. Does the site have different types or ranks of users such as members, moderators, or super-users? If you are unable to find an answer or if the question does not apply, leave the response blank.
   a. Yes
   b. No

4. Does the case have any of the following social network features? (Please indicate yes or no for each feature)
   a. "friend" or "follow" another user
   b. view "friends" or "followers" of other users
   c. subscribe to other users' updates or feeds
   d. join groups on the site with other users
   e. other social network functions

5. Do users on this site do any of the following things? (Please indicate yes or no for each)
    a. Recognize and/or label images
    b. Apply tags or labels
    b. Recognize and/or copy text
    d. ReCaptcha
    e. Rate or score things
    f. Vote
    g. Other very simple tasks

6. Does the site use any of the following features to filter content? (Please indicate yes or no for each feature)
    a. Recommendations
    b. Flagging
    c. Scoring or rating
    d. Other content filtering features
    e. No, none of the above
    f. Unable to answer this question

7. Please assess how many or how few opportunities exist for users to rate content (for example, scoring or ranking posts, comments, images, or movies):

    a. Very many opportunities to rate content
    b. Many opportunities to rate content
    c. Some opportunities to rate content
    d. Few Opportunities to rate content
    e. No opportunities to rate content
    f. Unable to answer this question

8. Please assess how many or how few opportunities exist for users to rate other users:
    a. Very many opportunities to rate other users
    b. Many opportunities to rate other users
    c. Some opportunities to rate other users
    d. Few Opportunities to rate other users
    e. No opportunities to rate other users
    f. Unable to answer this question

9. Does the site utilize any of the following content controls? (Please indicate yes or no for each control)
    a. Flagging of inappropriate content
    b. Account deletion
    c. Automated controls (e.g. limits on # contributions/hour)
    d. Instructions (on allowable behavior, etc)

10. Does this site make money from users in any of the following ways? (Please indicate yes or no for each method)
    a. Paid memberships
    b. Paywall for entry
    c. Donations
    d. Other source of revenue from users
    e. None of the above
    f. Unable to answer this question

11. What are the site's apparent sources of revenue? (Please indicate yes or no for each method)
    a. Advertising
    b. Memberships or subscriptions
    c. Donations (from users)
    d. Selling services
    e. Selling site-themed merchandise (swag)
    f. Selling goods
    g. Support from foundations, trusts, or charities
    h. Support from a parent company
    i. Support from some other outside organization (venture capital, private investors, etc.)
    j. Other revenue sources not in this list
    k. None, the site has no apparent revenue sources
    l. No information available to answer this question

12. Does this site make use of any of the following types of media? (Please indicate yes or no for each type of media)
    a. Text or writing
    b. Computer code (beyond what is used to run the website)
    c. Photographs or still images
    d. Video
    e. Audio
    f. Music
    g. Other forms of media

13. Please assess how many or how few of the site functions are accessible to non-registered users? (For example, posting content, reading content, viewing other user accounts)
    a. Very many functions accessible
    b. Many functions accessible
    c. Some functions accessible
    d. Few functions accessible
    e. No functions accessible
    f. Unable to answer this question.

**B. Legal**

14. Does the site use any of the following legal contracts? (Please indicate yes or no for each type of contract)
   a. Terms of Service
   b. Terms of Use
   c. Privacy Policy
   d. Copyright Agreement
   e. Intellectual Property Agreement
   f. End-User License Agreement (EULA)
   g. Some other contract
   h. Unable to answer this question

15. If possible, try to read a privacy policy, terms of service or license used on this site. Please assess how easy or difficult it is for a non-lawyer to read.

   Very difficult or impossible
   Fairly difficult
   Fairly easy
   Very easy
   Unable to answer this question

INSTRUCTIONS: The following questions ask about the use of copyright on the site. If it's not immediately apparent on the front page of the site, this information is often available in the site's terms of service; "contract," or "about" pages. It is possible that there may be multiple kinds of licenses in use.

16. Is there an explicit assertion of copyright on the site? If you are unable to find an answer or if the question does not apply, leave the response blank.
   yes
   no

17. Does the site as a whole use unmodified © or 'all rights reserved'? If you are unable to find an answer or if the question does not apply, leave the response blank.
   yes
   no

18. Does the site as a whole use Creative Commons (CC) licenses (www.creativecommons.org)? If you are unable to find an answer or if the question does not apply, leave the response blank.
   yes
   no

19. If the site as a whole uses CC licenses, indicate which options are available (Please indicate yes or no for each option):
    a. Attribution (BY)
    b. Non-commercial (NC)
    c. No-Derivatives (ND)
    d. Share-alike (SA)
    e. Public Domain (PD)

20. If the site as a whole uses other open licenses, indicate which options are used/available (Please indicate yes or no for each option):
    a. GNU Free Documentation License (GFDL)
    b. GNU Public License (GPL)
    c. Berkeley Software Distribution (BSD)

21. If the site as a whole uses a license not listed above, please name it here:_____

22. Does the site have distinct copyright licensing for user generated content, as separate from content uploaded by site administrators? If you are unable to find an answer or if the question does not apply, leave the response blank.
    yes
    no

23. Is user generated content covered by a single license? If you are unable to find an answer or if the question does not apply, leave the response blank.
    a. yes
    b. no

24. From the following options, please select who owns the rights over user-generated content:
    a. The site owner (individual or corporation)
    b. The user
    c. Other
    d. Unable to answer this question

25. Does the site make an unstructured dedication to the public domain (e.g. 'you can use this however you like')? If you are unable to find an answer or if the question does not apply, leave the response blank.
    a. yes
    b. no

## C. Assessment and Tags

26. Please indicate how strongly you agree or disagree with the following statement: "The site makes accurate claims about itself" (For example, information on the site's "about" page seems true)

Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree

27. If the site reports a number of users, how many are there? (Please enter a number and then select the corresponding time-frame)
     per year
     per month
     per week
     per day
     No time frame given
     Unable to answer this question

28. If the site reports a number of "active" users (as distinct from the previous question), how many are there? (Please enter a number and then select the corresponding time-frame) text field that can accept any number from 1-100,000,000
     per year
     per month
     per week
     per day
     No time frame given
     Unable to answer this question

29. If the site reports a number of readers or visitors, how many are there? (Please enter a number and then select the corresponding time-frame) text field that can accept any number from 1-100,000,000
     per year
     per month
     per week
     per day
     No time frame given
     Unable to answer this question

30. Please indicate how strongly you agree or disagree with the following statement: "this site features thriving activity" (For example, regular contributions from users, recent feature updates, etc)
     Strongly Agree
     Agree
     Neutral
     Disagree
     Strongly Disagree

31. Please assess how easy or how difficult it is to use this site:
        Very easy to use
        Easy to use
        Difficult to use
        Very difficult to use

32. Does the site fall into any of the following categories? If you can think of a category that might be relevant, please add it at the end in the space provided. (Please indicate yes or no for each category).
        a. Media (for example: music, images, photos, text, or movies)
        b. Social networks or groups
        c. Gaming or role-play
        d. Discussion (for example, blogs, forums, message boards, chat)
        e. Design or production (for example, of software, t-shirts, or shoes)
        f. Distributed resources (for example, peer-to-peer networks, computing)
        g. Information pools (activities that create shared information like encyclopedias)
        h. Crowdsourcing and outsourcing
        i. Social filtering (for example, of news, books, movies, or music)
        j. Other [text box]

## D. Organization and Leadership

33. Organizational type: This site is owned or operated by: (Please indicate yes or no for each of the following options):
        a. For-profit organization or company
        b. Non-profit or foundation
        c. University
        d. Government
        e. Some other formal organization (none of the above or unclear)
        f. No formal organization
        g. Unable to answer this question

34. Can you find out who is in charge of this site or community?
        Yes, I found this information
        Maybe; I think I found this information
        No, I cannot find this information

For the following questions, please indicate how common or unusual each action is on the site:

35. Users elect community leaders:
Very Common --- Somewhat Common --- Somewhat Unusual --- Very Unusual

36. Users vote in decisions about the site (for example, votes about content, rules, or changes to the site)

Very Common --- Somewhat Common --- Somewhat Unusual --- Very Unusual

37. Users elect company or project leaders:
Very Common --- Somewhat Common --- Somewhat Unusual --- Very Unusual

38. Users vote in decisions about the company or project (for example, votes about strategy, hiring or firing, or organization changes):
Very Common --- Somewhat Common --- Somewhat Unusual --- Very Unusual

39. Users deliberate in decisions about the site (for example, decisions about strategy, hiring or firing, or organization changes):
Very Common --- Somewhat Common --- Somewhat Unusual --- Very Unusual

40. Users deliberate in decisions about the site (for example, decisions about content, rules, or changes to the site):
Very Common --- Somewhat Common --- Somewhat Unusual --- Very Unusual

Instructions: For these next questions, you can usually find the information to answer the following question in the site's F.A.Q., "about," or "meta" pages. Select the answers that are most consistent with these or other authoritative sources from the site. If you are unable to find an answer or if the question does not apply, leave the response blank).

Who has authority over the site?

41. An individual or small group (not a company)?
   yes/no
42. A company (managerial/corporate)?
   yes/no
43. A group of elected community representatives or super-users (virtue republic)?
   yes/no
44. Community members and leaders share authority (distributed governance)
   yes/no
45. If authority is distributed, are final decisions about governance still made or approved by a centralized body of some kind (if yes, core/periphery)?
   yes/no
46. If authority is distributed, do governance decisions require the consensus or approval of all users/participants (if yes, direct democracy)?
   yes/no
47. If authority is distributed, do governance decisions happen through negotiations among the people involved/interested in a particular issue (constitutional anarchy)?
   yes/no

## E. Norms and Enforcement

48. Does the site ask users to follow certain standards of acceptable behavior? (If you are unable to find an answer or if the question does not apply, leave the response blank).
  yes/no

49. Please indicate how frequently site users can enforce site rules or norms.
  Very frequently
  Frequently
  Sometimes
  Rarely
  Very rarely or not at all

50. How strongly do you agree or disagree with the following statement: "When a conflict happens on this site, users settle it by themselves."
  Strongly Agree
  Agree
  Disagree
  Strongly Disagree

51. Does the site resolve disputes in any of the following ways (either between users or between administrators and users)?

52. How is behavior monitored on the site? (More than one option may be applicable, so please indicate yes or no for each of the following options).
  a. Site leaders or administrators monitor users
  b. Users monitor each other (through flagging, reporting, or other means).
  c. Automated bots, filters, or similar means
  d. Some other form of monitoring
  e. No visible forms of monitoring
  f. Unable to answer this question

Instructions: For the following questions, please indicate yes or no for each potential answer option.

53. Do site leaders or administrators reward users for good contributions...
  ...with money  (yes/no)
  ...with public praise like reputation scores or stars (yes/no)
  ...with special responsibilities or privileges (yes/no)

54. Do users reward each other for good contributions...
  ...with money  (yes/no)
  ...with public praise like reputation scores or stars (yes/no)
  ...with special responsibilities or privileges (yes/no)

55. Do site leaders or administrators punish users for bad contributions...
        ...with fines    (yes/no)
        ...with public shame   (yes/no)
        ...with banning or blocking accounts (yes/no)

56. Do users punish each other for bad contributions...
        ...with fines    (yes/no)
        ...with public shame   (yes/no)
        ...with banning or blocking accounts (yes/no)

## F. User Behavior

57. Please assess the level of user-to-user interaction on this site (for example, how much do users communicate with each other, exchange ideas, and evaluate each other's actions?):
        Very high level of user-to-user interaction
        Above average level of user-to-user interaction
        Average level of user-to-user interaction
        Below average level of user-to-user interaction
        Low level of user-to-user interaction
        Unable to answer this question

58. Users of the site can view each other's contributions or actions.
        Strongly Agree
        Agree
        Neutral
        Disagree
        Strongly Disagree

59. Users of this site can respond to each other's contributions or actions.
        Strongly Agree
        Agree
        Neutral
        Disagree
        Strongly Disagree

60. Please indicate how strongly you agree or disagree with the following statement: "Users of this site often give each other praise or symbolic rewards (stars, reputation points)"
        Strongly Agree
        Agree
        Neutral
        Disagree
        Strongly Disagree

61. How common or unusual are each of the following activities among users of this site? Indicate your assessment on a scale of "very common" to "very unusual":
   a. Creating reports or records of events
   b. Analyzing information, issues, or ideas
   c. Cultivating specialized skills
   d. Innovation or invention
   e. Assessing the relevance or importance of something
   f. Assessing the quality of something
   g. Predicting the outcome of an event or process
   h. Creating cultural (for example, artistic or religious) works

62. Please indicate how confident you are that the following statement about the site is TRUE: "Users of this site behave according to informal rules or codes"
   Very confident
   Somewhat confident
   Not very confident
   Not at all confident

## G. Motivations

Content: Please assess how important or unimportant each of the following kinds of content is for attracting people to the site:

63. Written text
   Very Important
   Somewhat Important
   Somewhat Unimportant
   Very unimportant

64. Images
   Very Important
   Somewhat Important
   Somewhat Unimportant
   Very unimportant

65. Video
   Very Important
   Somewhat Important
   Somewhat Unimportant
   Very unimportant

66. Text, images, or video created by professionals
   Very Important
   Somewhat Important
   Somewhat Unimportant

Very unimportant

67. Text, images, or video created by site owners
     Very Important
     Somewhat Important
     Somewhat Unimportant
     Very unimportant

68. Text, images, or video created by other users of the site
     Very Important
     Somewhat Important
     Somewhat Unimportant
     Very unimportant

Participation incentives: For each of the following incentives, indicate whether it is present or absent on the site. If you are unable to find an answer, you should leave that item blank.

(Choice of present or absent for each of these questions)

69. Solving practical problems
70. Having fun
71. Being entertained
72. Socializing with friends
73. Giving or receiving emotional support
74. Creative self-expression
75. Learning
76. Learning specialized skills
77. Doing "the right thing'" (fulfilling an ethical or moral ideal)
78. Earning admiration of others
79. Promoting a cause
80. Interacting with like-minded people
81. Expressing your personal identity
82. Achieving shared goals together with others
83. Collaborating offline with others
84. Promoting yourself or your work
85. Earning access to exclusive parts of the site
86. Winning a contest
87. Winning prizes (not money)
88. Solving puzzles
89. Making professional contacts
90. Getting a job
91. Doing your job better
92. Making money
93. Earning virtual currency (such as points)
94. Promoting a company

Motives for participation:  Please how much the site emphasizes each motive or not:

Answer choices for all of the following questions:

Emphasizes a lot
Emphasizes somewhat
Emphasizes a little
Does not emphasize at all

95. Solving practical problems
96. Having fun
97. Being entertained
98. Socializing with friends
99. Giving or receiving emotional support
100. Creative self-expression
101. Learning
102. Learning specialized skills
103. Doing "the right thing"' (fulfilling an ethical or moral ideal)
104. Earning admiration of others
105. Promoting a cause
106. Interacting with like-minded people
107. Expressing your personal identity
108. Achieving shared goals together with others
109. Collaborating offline with others
110. Promoting yourself or your work
111. Earning access to exclusive parts of the site
112. Winning a contest
113. Winning prizes (not money)
114. Solving puzzles
115. Making professional contacts
116. Getting a job
117. Doing your job better
118. Making money
119. Earning virtual currency (such as points)
120. Promoting a company

Reasons for using the site: Please indicate how important or unimportant each reason is for the site users:

Answer choices for all of the following questions:
Very important
Somewhat important
Somewhat unimportant
Totally unimportant

121. Solving practical problems
122. Having fun
123. Being entertained
124. Socializing with friends
125. Giving or receiving emotional support
126. Creative self-expression
127. Learning
128. Learning specialized skills
129. Doing "the right thing"' (fulfilling an ethical or moral ideal)
130. Earning admiration of others
131. Promoting a cause
132. Interacting with like-minded people
133. Expressing your personal identity
134. Achieving shared goals together with others
135. Collaborating offline with others
136. Promoting yourself or your work
137. Earning access to exclusive parts of the site
138. Winning a contest
139. Winning prizes (not money)
140. Solving puzzles
141. Making professional contacts
142. Getting a job
143. Doing your job better
144. Making money
145. Earning virtual currency (such as points)
146. Promoting a company

## H. Tasks

Instructions: Who decides how users participate in this site? For example: if users upload photos, who decides how many photos? Who decides the content of the photos? Who decides how the photos should be labeled? Please take these questions into account and indicate how much each of the following groups define user participation or not:

147. Site administrators or leaders:
     Completely define user participation
     Somewhat define user participation
     Rarely define user participation
     Do not define user participation at all

148. Individual users:
     Completely define user participation
     Somewhat define user participation
     Rarely define user participation
     Do not define user participation at all

149. Other users:
Completely define user participation
Somewhat define user participation
Rarely define user participation
Do not define user participation at all

How strongly do you agree or disagree with the following statements about the site:

150. Some users of this site do specialized tasks
Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree

151. Some users of this site do simple tasks
Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree

The following questions refer to different types of information flow. For each type, we provide a brief description and then ask a couple of questions.

TYPE I: The flow of information from many-to-one-to-many (For example: a site that collects many users' opinions and then shares the aggregated results)

152. How common or unusual is this kind of information flow on this site?
Very common
Somewhat common
Somewhat unusual
Very unusual

153. Is this kind of information flow important or unimportant to the site?
Very important
Somewhat important
Neither important nor unimportant
Somewhat unimportant
Very unimportant

TYPE II: The flow of information from some-to-many (example: a small group blog or organization distributing information to many people)

154. How common or unusual is this kind of information flow on this site?

Very common
Somewhat common
Somewhat unusual
Very unusual

155.    Is this kind of information flow important or unimportant to the site?
Very important
Somewhat important
Neither important nor unimportant
Somewhat unimportant
Very unimportant

TYPE III: The flow of information from some-to-some (example: a community or site where users interact with each other in small groups)

156.    How common or unusual is this kind of information flow on this site?
Very common
Somewhat common
Somewhat unusual
Very unusual

157.    Is this kind of information flow important or unimportant to the site?
Very important
Somewhat important
Neither important nor unimportant
Somewhat unimportant
Very unimportant

TYPE IV: The flow of information from many-to-many (example: a large network where individuals and groups exchange information in a very decentralized manner)

158.    How common or unusual is this kind of information flow on this site?
Very common
Somewhat common
Somewhat unusual
Very unusual

159.    Is this kind of information flow important or unimportant to the site?
Very important
Somewhat important
Neither important nor unimportant
Somewhat unimportant
Very unimportant

Instructions: The next questions ask whether the contributions to the site are interdependent or not. Please state how strongly you agree or disagree with each statement:

160.      An individual can make a contribution to the site which can stand alone (like a photograph or a blog post)
Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree

161.      Many individuals contribute to the site and their contributions are interdependent in some way (like the words in a sentence, or the notes in a song)
Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree

162.      If you wanted to participate in this site, about how much time would you need to do it for the first time?
a. Less than 5 minutes
b. More than 5 minutes
c. I cannot find this information

163.      If you wanted to contribute to this site on a regular basis, about how much time would a small contribution require?
a. Less than 5 minutes
b. More than 5 minutes
c. I cannot find this information

164.      Does the case produce any of the following work outputs ? : (Check all that apply)
Unskilled labor
Semi-skilled labor
Skilled labor

By participating, do users generate any of the following resources? (y/n)
Computing resources
File sharing

166.      Does the site distribute any of the following resources? (y/n)
User investments (money)
User donations (money)

Material goods (products)

167.      Does the site ask users to contribute any of the following resources? (y/n)
Physical stuff (goods)
Hard drive space
Computing power

User location: Indicate how strongly you agree or disagree with the following statements:

168.      "This site functions well because users are in particular locations."
Strongly Agree
Agree
Disagree
Strongly Disagree

169.      "This site functions well because users are in the right place at the right time."
Strongly Agree
Agree
Disagree
Strongly Disagree

# OCR Progress Report Annex 2

## Phase II Study Abstracts

*Title:* "Collaborative News Reporting and Citizen Journalism Sites"
*Author:* Yael Granot
*Study Type:* Domain-specific; exploratory; internal
*Abstract:*

This study examines sites engaged in the collaborative production of citizen journalism or participatory journalism. Each of the cases involve people worldwide volunteering their own written articles, photos or videos in response to current world events. Some of the cases include paid staff writers or aggregated elements from other sources, but all have some distinct component of volunteer news contribution. Comparison across the different sites revealed that motivational appeals to ideals of journalistic excellence were especially widespread, suggesting an area for future research within this domain.

*Title:* "Hospitality Websites: Design Levers for Cooperation"
*Author:* Roxanna Myhrum
*Study Type:* Domain-specific; exploratory; internal
*Abstract:*

The sites in this study facilitate "hospitality exchange." This term encompasses international cultural exchange and the formation of friendships between travelers, as well as the provisioning of free accommodation to travelers by participants residing in a host country. All of the sites in this study are backed by non-profit entities or operate in a not-for-profit manner. All appear to have at least some current user activity and up-to-date profile information. The sites share several overlapping cooperative problems: Participants in the community need to be able to trust one another to a) behave appropriately as travelers, b) to provide reliable and safe accommodation as hosts, c) to not misuse personal information revealed to the site or community, and d) not to exploit their personal interaction once it is no longer subject to direct enforcement by the site. Additionally, three sites faced the challenge of distributing and coordinating the effective management of the site's operations amongst a large team of volunteers.

*Title:* "Virtual Airlines"
*Author:* Anna Y. Kim
*Study Type:* Domain-specific; exploratory; internal

*Abstract:*

This study compares several Virtual Airline communities as examples of cooperative communities online. Virtual airlines mirror real world airlines and use flight simulation to model the operations of an airline. With widespread use of broadband internet connections, virtual pilots can connect to servers allowing them to share the same virtual airspace.  The website www.avsim.com lists over 300 "Passenger Oriented Vas."  Each airline has its own Web Page, some of which are extraordinarily detailed, realistic and sophisticated.   In general, virtual airlines seek to provide compelling and realistic experiences similar to operations inside a real airline. Users sign up for free and start off by "flying" the smallest aircraft.  As he/she builds hours and skills, she may earn promotions and build their reputation within the airline.

*Title:* "Collaborative Fashion Sites"
*Authors:* Dharmishta Rood and Anna Kim
*Study Type:* Domain-specific; exploratory; internal
*Abstract:*

This Phase II report examines collaborative websites that deal with fashion and style. For the purposes of this study, we are restricting the the domain of fashion and style websites to only those that in some way facilitate the creation of clothing and personal adornment accessories (jewelry, handbags, shoes, etc.). The sites themselves range from internationally renowned collaborative t-shirt websites, to small sites that allow individuals to alter clothing patterns. These sites were selected to help understand the ways online collaboration can facilitate the production of goods to be consumed offline. We're interested in the motivations and incentives used within these communities as well as the ways community members interact with one another's creative work. These sites were chosen carefully to encompass a genre of online fashion production that both challenge and re-create traditional models of large-scale production of fashion goods.

*Title:* "Collaborative Entrepreneurship and Business Communities"
*Author:* Leah Belsky
*Study Type:* Domain-specific; exploratory; internal
*Abstract:*

This domain includes sites that enable entrepreneurs and professionals to meet, connect, give advice, and provide services. The site encompasses the massively popular "LinkedIn" as well as smaller, niche sites geared towards entrepreneurs. These sites are solving a few main problems. First, they are allowing entrepreneurs to find and meet one another. After establishing the social network, the sites then allow professionals to share experiences, form groups, get advice about forming companies, and contract for services. A major question raised by this study whether creation and formation of social networks – with concomitant advice and support functions – can meaningfully be considered

collaborative. In response to this concern, I propose decomposing our definition of collaboration so that we pull apart the "weaker" forms of collaboration – like social networking – from more meaningful forms of social production. I've taken initial crack at this using the survey codebook to distinguish between what we might call "collaboration 1.0" and "collaboration 2.0" at the end of this report.

*Title:* "Bad Guys: Illegitimate and Illegal Collaboration Sites"
*Author:* Ayelet Oz
*Study Type:* Domain-specific; exploratory; internal
*Abstract:*

The "bad guys" domain focuses on what might be defined as collaboration for illegitimate goals. In structuring the participating sites, I have tried to broaden this definition to include various illegal, violent, or destructive groups. The sites chosen could be categorized into four groups – hacking sites, racist sites, anarchist sites and sites that aim to support illegitimate behaviors (pedophiles and anorectic girls). All of the sites chosen share a common dilemma between the need to be observable enough to reach their audience yet hidden enough not to be shut down by enforcement authorities or attacked by the general public. Several shared characteristics emerge across the cases in the study: (1) they tend to focus on off-line activities; (2) they emphasize of shared goals; (3) news and information make up the central outputs of most of the sites; (4) the groups demonstrate informed decision-making regarding licensing and privacy. These characteristics reflect the reality that leaders of these sites rely on and enforce prior shared normative goals or political identities of their user community, rather than on generating new motivations for participation.

*Title:* "The Daily Kos: A Case Study of a Large-Scale Collaborative Political Blog"
*Author:* Aaron Shaw
*Study Type:* Domain-specific; qualitative case study; exploratory; internal
*Abstract:*

Since its creation, The Daily Kos (http://www.dailykos.com) has evolved into the most heavily trafficked, collaborative political blog on the Internet. Generally characterized as radically left-wing, the site has become the pre-eminent symbol of the so-called "Netroots" movement as well as the broader explosion of online political discourse in the United States. While many online news outlets can match the quantity of traffic on Daily Kos, none have harnessed user-generated content to a comparable extent. The extensive use of distributed information production and filtering on Daily Kos makes it interesting case for political communications research and studies of large-scale networked cooperation. This paper provides an historical overview of the site, with an emphasis on the structures and practices that facilitate the large-scale information production and filtering that are the community's hallmark.

*Title:* A Tale of Two Blogospheres: Discursive Practices and the Left and Right
*Authors:* Yochai Benkler, Aaron Shaw, and Victoria Stodden
*Study Type:* Domain-specific; observation survey; content analysis; collaboration
*Abstract:*

We compared the technologies, practices, and discursive structures of the left and right wings of the top 155 political blogs in the United States in the summer of 2008. Contrary to earlier claims that both sides of the blogosphere were relatively symmetric, we found that the left and the right differed significantly along several dimensions pertinent to the structure of networked discourse. Despite substantial overlap, we find that the left and right adopted significantly different technological approaches, with the left using platforms and plugins that enhance articipation and discussion to a much greater extent than the right. In all, the right wing of the blogosphere was more likely to reflect a more individualistic and hierarchical style: that is, sole authorship with active commentary by users/readers playing a more peripheral role. In contrast, the left of the political blogosphere exhibited greater collaborative affordances and more egalitarian and participatory discursive practices. (Note: We are in the process of revising this article and plan to issue it as a Berkman Center Working Paper later in June, 2009. We include it here primarily because it included an early version of the questions and content analysis techniques which we are also using for Phase I of our large-scale observation survey.)

*Title:* "Driving the Hive: An Experiment in Motivation and Task Design in a Distributed Labor Market"
*Authors:* Daniel Chen, John Horton, Aaron Shaw and Yochai Benkler
*Study Type:* Experimental; cross-domain; collaboration
*Abstract:*

This study examines the effect of different motivational framings on worker performance in Amazon Mechanical Turk (AMT), a distributed online labor market. Drawing on previous literature from several disciplines, we conduct a controlled experiment with AMT workers in which workers are exposed to one of sixteen treatments before completing a qualitative analysis task in which they answer objective questions about a website. We then judge worker performance against "gold standard" responses to the same questions provided by expert raters. (Note: Data collection for this study is currently underway and will be completed during June and July, 2009. The study is also serving as a preliminary test of the ScriptGen tool and a pilot study that will determine how we implement the Distributed Observation Survey for Phase I above.)

*Title:* "Wikipedia as a Window on the World? Notability and Bias in the World's Largest Free Encyclopedia"
*Authors:* Benjamin Mako Hill and Aaron Shaw

*Study Type:* Analytical; domain-specific; collaboration
*Abstract:*

How effectively does Wikipedia document the world? In this paper, we analyze the extent of Wikipedia's coverage in order to better understand the implications of its criteria for inclusion and its editorial process for knowledge production. Specifically, we construct quantitative models to compare Wikipedia's coverage against exhaustive data-sets that document particular domains of activity, such as films, open source software projects, firms, and non-profit organizations. The results provide insights on just how inclusive "The World's Largest Free Encyclopedia" actually is, as well as a set of tested strategies for measuring bias within collectively edited textual corpora.

*Title:* "Collective Intelligence Sampling: Mining Wikipedia to Discover a Sample"
*Authors:* Aaron Shaw and Hal Roberts
*Study Type:* Analytical, domain-specific, methodology; collaboration
*Abstract:*

Can the contents of Wikipedia serve as a robust sampling frame for poorly defined or hidden populations of websites? The method of Collective Intelligence Sampling we describe and test in this paper overcomes the difficulties of sampling online phenomena by leveraging our existing knowledge of a particular domain together with computational data-mining in order to discover a large sample of sites that meet well-defined criteria. Adapting a toolkit developed by David Milne (Milne and Witten 2008), we apply algorithmic filtering and text-mining techniques to the English language Wikipedia (en:WP) in order to extract a sample of articles with high semantic relatedness and network proximity to a set of pre-validated keyword terms. We then review the results of the automated technique and use human coders to validate whether the results meet criteria for inclusion in our population or not.

# OCR Progress Report Annex 3

## Hospitality Websites: Design Levers For Cooperation

**Roxanna Myhrum**
**April, 2009**

### ABSTRACT

The sites in this study facilitate "hospitality exchange." This term encompasses international cultural exchange and the formation of friendships between travelers, as well as the provisioning of free accommodation to travelers by participants residing in a host country. All of the sites in this study are backed by non-profit entities or operate in a not-for-profit manner. All appear to have at least some current user activity and up-to-date profile information. The sites share several overlapping cooperative problems: Participants in the community need to be able to trust one another to a) behave appropriately as travelers, b) to provide reliable and safe accommodation as hosts, c) to not misuse personal information revealed to the site or community, and d) not to exploit their personal interaction once it is no longer subject to direct enforcement by the site. Additionally, three sites faced the challenge of distributing and coordinating the effective management of the site's operations amongst a large team of volunteers.

### Site Selection Criteria

The sites in this study facilitate "hospitality exchange." This term encompasses international cultural exchange and the formation of friendships between travelers, as well as the provisioning of free accommodation to travelers by participants residing in a host country. All of the sites in this study are backed by non-profit entities or operate in a not-for-profit manner. All appear to have at least some current user activity and up-to-date profile information. The sites in this study were gathered through Google and Wikipedia searches augmented by reviews in travel publications. In some cases, sites also contain information about others in the study, occasionally defining themselves in contrast to their competitors.

### Cooperative Problems

Participants in the community need to be able to trust one another to a) behave appropriately as travelers, b) to provide reliable and safe accommodation as hosts, c) to not misuse personal information revealed to the site or community, and d) not to exploit their personal interaction once it is no longer subject to direct enforcement by the site. Additionally, three sites faced the challenge of distributing and coordinating the effective management of the site's operations amongst a large team of volunteers.

**The Sites**

**Global Freeloaders** was founded in 2000 by Adam Staines, a 20-something Australian who loved traveling and who had personally benefited from staying with locals during his backpacking adventures. He currently appears to be the sole administrator of the site, and requests donations from members to pay for site maintenance expenses. Adam is also a user of the site and identifies both as a "freeloader" and a "traveler." He frequently offers his own experiences as the ideal-type example of how to behave on the site. The dual missions of seeking free accommodation and being an open-minded world explorer are maintained throughout the site and help to define its culture. They also resonate on the site's "reference" pages where users can comment on both the "character" of their hosts (human friendliness, awesomeness, etc.) and the "accommodations" provided (did you get what you paid for).

Registration on this site proceeded without a hitch, and the site does not appear to be buggy. However, the most recent communications from the site founder are dated several years ago. There is a page that breaks down membership by country. Adding these numbers indicates that there are approximately 66,952 current members with active profiles, concentrated primarily in English-speaking countries (US, 21,072; Australia 7128; UK, 5145; Canada, 4664; and Germany, 3184). 212 countries are listed.

Due to Adam's (jokingly self-proclaimed linguistic) limitations, the site is only available in English. Issues related to gender are not conspicuously addressed anywhere on the site, although hosts can state an open-ended preference for the gender (or age, or values) of their guests, as well as any other exclusionary preferences they would like.

Users of the site are strongly expected to balance their traveling with hosting. The site facilitates batch requests for housing based on geography, and then encourages users to communicate over e-mail or chat until they feel comfortable offering their personal address details. The obligation is skewed strongly towards the "freeloader" to prove his or her worthiness of receiving free accommodation. Users can leave recommendations ("references") about each other (both as host and guest).

The site does not use pictures on its user profiles. It has the most open-ended format for providing profile information (two large text boxes) where users can describe themselves and what the want/have to offer as hosts. It does not have an open communications forum, although testimonials have periodically been collected and posted, and communication with the site administrator is possible.

**Hospitality Club** was founded in 2000 by Veit Kuehne, a 22 year old living in Germany. Prior to founding the site, he had been an exchange student and a member of other organizations with international memberships (Mensa). He desired to expand on these experiences and hospitality exchange. Viet also positions HC as taking the values and practices of the offline SERVAS network onto the Internet. Veit is still the voice of some pages of the site, is still registered as a user, and appears to be promoting the site through word of mouth. However, the club's numerous volunteers appear to direct the site by

implementing content and making decisions about members. Volunteers are recruited to the cause throughout their registration process. Viet's leadership seems to have been the cause of members migrating away from Hospitality club to other sites.

The site utilizes some google ads to "cover webhosting costs," although none were observed on the site. A mechanism for donations was not observed. The site uses friendly, colloquial language (with occasional imperfect English). It is dominated by text, and has several pages that have not been updated for several years, although there were several forum threads from 2009. There is some disagreement in the total number of members reported on the site (there are different numbers on different pages), but it seems to be somewhere between 380,000 and 500,000 The site is available in many languages aside from English, and users are permitted to write in whatever language they want on their profile, although English is named as the most popular language. Profiles allow users to specify the gender of travelers they desire to host, but no official statement on gender and travel was observed on the site. .

As might be expected from the term "club," the site promotes solidarity amongst its members and encourages fidelity to the club itself. This is achieved through consistent emphasis on the site's mission to create a "world wide web of friendly people," bolstered by rules mandating compliance with the 'Hospitality Club Spirit." Expressions by HC's users indicate that they are very passionate about the site's cause and the traveler/backpacker/hitchhiker lifestyle. There is observable tension on the site (observed in some of the general writing and in the forum) between its desire to expand on and recruit more members globally, and its desire to maintain and enforce the strong ethic in use by other similar offline clubs with higher barriers to entry or in-person screenings. This remains an issue under discussion by the site's members and volunteers.

Out of all the sites in the sample, this one has the strongest requirement extending offline to the face-to-face interaction: members are required to exchange passport numbers up front, and a traveler is required to show his/her passport to the host.

*NOTE: The site and some server trouble initially when I tried to sign up, and because volunteers review the applications mine is still under consideration. According to some sites critical of Hospitality club, this process may take months.*

**BeWelcome** and its associated BeVolunteer network were founded in February 2007 by former volunteers from Hospitality Club who were dissatisfied with the management of that site's content and volunteer community. They also desired that the site utilize an opensource platform, and so is built on BW-rox. The project also attracted volunteers who were dissatisfied with couchsurfing (example: the opencouchsurfing.org website). The site is still in its beta phase, and statistical graphics show approximately 6250 "friendly active members."

The site's administration is Euro-centric (The official non-profit volunteer organization is incorporated in France). The site currently has several language mirrors, with ambitions

for volunteers to carry out translation into additional languages.  English is the required lingua franca of the volunteers.

The site utilizes profiles with commenting ability to coordinate hospitality exchange.  Members are encouraged to investigate profiles until they feel comfortable around another user.

The core of this site's activity appears to be located in its volunteer network rather than in the sign up member service (users seem to have more robust reputations in this site than in the reputation system in their profiles).  Hence, those users who are drawn to the volunteer network and its strong values (anti centralized authority, open source, etc.) seem to be more committed and active on the site.  As the volunteer community reaches a critical mass, they may utilize the hospitality functions of the site more and produce more activity in this area of the site.

**Couchsurfing** was founded in 2004 by Casey Fenton, Sebastien Giao Le Tuan, Daniel M. Hoffer, and Leonardo Bassani da Silveira, all of whom were young, entrepreneurial travelers/backpackers and who shared a desire to institutionalize and scale up a web-based service for travel accommodation.  However, they state that their core values are not about saving money, but rather about idealism, cultural exchange, making new friends, and faith in others.  The site's current mission is *"CouchSurfing seeks to internationally network people and places, create educational exchanges, raise collective consciousness, spread tolerance, and facilitate cultural understanding."* The site currently reports approximately 1,088,166 members in 231 countries.

The site is run by Couchsurfing International, a non-profit organization.  The site has revenue from selling themed products, and from its identity verification service.  It also accepts donations from users.  It does not display advertising. The site runs a large and highly coordinated volunteer network, coordinated through two offices: one in California, and one that floats around the world (and offers free housing to volunteers located there).  Volunteers manage all of the site's major operations.  There is evidence that some volunteers have felt they did not have enough say in the organization's governance and left to participate in other organizations.

The site proactively and conspicuously addresses questions of safety.  It runs an optional verification system for user identity, and makes most of its revenue from this service (costs are variable by country). In addition to on-profile recommendations, it also has a "vouching" feature with substantial rules which ties the reputation system to the site's social network.  Additionally, safety tips are offered, including a special section of advice for female travelers.

User profiles have extensive features for humanizing users, and request pictures of a user, his/her house, and his/her travel experience are all welcomed.  Upon registering for the site, a volunteer ambassador contacts you via e-mail and requests that you fill out your profile more thoroughly.

"Couch surfer" has come to be used to identify the users of the site. Users have coordinated offline interaction among couchsurfers in their region. These events are more of a celebration of shared values than an instrumental usage of the site.

**Stay4Free**

Stay4Free is a home exchange network that, in contrast to the other sites, emphasizes accommodation exchange, including house swapping, over meaningful in-person user interaction (although it does allow for the possibility that users will interact and make friends while they are coordinating their plans or offering accommodation).

The site is very opaque about its operations, and very little information was available about its founding, governance, or staffing. According to the homepage, the site is a non-profit organization that was founded by a "group of world travelers" in partnership with PLANT, their web host, which is based in the Netherlands.

Registration was very simple (there is no requirement to enter extensive personal information) and proceeded smoothly. Searches did return several user profiles (from large European cities) with what appeared to be active e-mail addresses. Overall, the site appears to be more of a web application for coordinating a non-monetary market in international accommodation.

The site has a button requesting donations, and the terms of service indicate that targeted advertising is used (although none was observed).

**Design Levers in Use**

| Design Element | How used on Hospitality Sites |
|---|---|
| Communication | All members need to communicate before they interact in real life, both to coordinate their plans and to form trust. Variations: on site/e-mail communication, user forums, communication with administrator, communication with volunteer network, periodic gatherings of site users. |
| Empathy/Humanization | All users need to at minimum verify their identity as a real person, which is often done in a regimented technical system on the sites. Beyond that, sites use story telling, profile pictures and travel photos, open-ended conversation (in forums and on profiles) to humanize the participants. Significant variation in this feature is displayed across sites (couchsurfing addresses this the most, Stay4Free the least). |
| Solidarity | Solidarity can be formed around a) the site itself, b) the ideal-type user identity (ex: "travelers" or "idealists") c) other pre-determined identifiers (language, nationality, gender) d) interest group and e) shared experience. The sites display variation in how they promote solidarity along these axes. Many sites have developed an identification for users that corresponds with the site name ("Global freeloaders," "couchsurfers") or with |

| | various levels of site participation ("ambassadors," "volunteers"). |
|---|---|
| Fairness | There are multiple scales of fairness on these site: some emphasize balance of hosting and receiving accommodation; others emphasize a balance between using the site and contributing to its maintenance through donations or volunteering. |
| Norms | Most of these sites have strong values articulated for their members. In all cases except Stay4Free (and possibly there as well), these norms relate to the values of the sites' founders. In general, the sites articulate clear procedures for initiating communication and carrying out offline interaction (also strongly tied to idealistic values), and for evaluating the success of that interaction back on the site. |
| Trust | In general, sites create a background of trust in the ideals of the project that primes users to a certain mindset and way of behaving before they transact with individuals. While many sites have rule enforcement mechanisms for defectors (and criminals), the emphasis in most cases is on recruiting trustworthy people who share values. In no case is transacting fully automated; all interaction between users is subject to their own discretionary feeling of comfort. |
| Efficacy | Efficacy in this domain refers both to measuring individual experiences and to measuring the success of the site itself (and accordingly, its idealistic mission). All but one site allows for reviewing of past experiences; these sites also use extensive statistical graphics or charts to show the total number of users and other information. All sites allow users to modify their profiles, and emphasize the relationship between a complete profile and effective communication. |
| Punishment/reward | The sites do not appear to utilize incentives other than the valuable travel experience facilitated by the site and the satisfaction of users' intrinsic motivation to help perpetuate the site's community. For example, users' efforts to recruit new users are framed as helping to accomplish the shared goals of the community and make the world a better place for all, rather than as a recruitment competition. Punishment on the sites has various forms, and ranges from the issuing negative reviews by users to expulsion from the site by administrators. In three cases, volunteers undertake efforts to monitor profile content and/or user behavior to ensure compliance with site norms. |
| Crowding out | The greatest tension between design levers that I observed resulted from members having issues with the governance structure of a site. In these cases, a clear set of norms derived from the founder clashed with what participants believed to be effective operation. In only one case was tension between the |

| | |
|---|---|
| | "free accommodation" and "cultural exchange" ideals observed; in most situations, sites appear to mitigate the selfish search for free accommodation with a values-based "traveler" ethic. |
| Transparency (reputation) | Transparency in the user community is required to facilitate repeated transactions by particular hosts or guests. A person's *character* reputation was transportable across both sides of the hospitality transaction (host and guest), whereas his or her *accommodation* reputation only applied to the quality of his or her lodgings. Reputations are associated with a users' profile or usernames through reviews and ranking assessments. The sites displayed variation in the technology and communication mode used to collect and disclose reputations.<br>Transparency about the organization or the individual behind the site varied significantly. In most cases, sites strongly emphasized their willingness to be transparent to members about their operations as a shared value of the site.<br>Many sites keep significant portions of their content opaque to visitors until they become members. |
| Cost | Cost of participation equals the time and effort spent to compose an appealing profile, to signal trustworthiness to other members, and to coordinate travel intent. This cost is framed as relatively low in all cases, as it is measured against an equivalent (expensive) process used to secure accommodation through traditional means. All services are free, although some request donations or volunteer time. |
| Exit/entry | All communities require a registration process. While site tours are not mandated, most potential users are able to acquaint themselves with the norms and values of the site before entering, and 4 sites provide testimonials by current users. Exiting the community seems to happen by disinterest/abandonment in most cases. Most sites identified managing inactive users and deleting accounts as a task. |
| Leadership/asymmetric contribution | As mentioned, the founders are significant fixtures on all but one site, and they lead the community by example. 3 sites maintain extensive volunteer communities with their own internal management hierarchies who do the work needed to maintain the site and community. These volunteer communities display varying modes of governance. It is clear on most sites that some users are more active travelers than others, and accordingly have deeper reputations. In general, all users are empowered to promote the site to their social networks and recruit new users through word of mouth. |

**Questions for Follow up:**

- What patterns emerge from the use of various design levers? *Note: What method would be best for evaluating this? Ideally I would develop a matrix that codes the sites based on certain sub-features of this domain would be useful (ex: can users also act as volunteers, yes/no? Is there a delay mandated before you have full access to the profiles on the site, yes/no).*
- How does the competition between sites influence the user experience? How does it influence the shaping and evolution of site norms?
- How does each site resolve conflicts about governance? What implications does this have for the core activity of the site? How can this be understood in terms of "crowding out?"
- What are the issues that arise as the site grows in membership and diversity?
- How do users perceive the privacy and security measures on the site?
- Does intensity of agreement with the site's values have an influence on intensity of a user's participation?
- How does offline word of mouth promotion take place? How do users promote the site using their social networks? What are the ways that users are initially drawn to the site (free accommodation, a desire to host, friend recommendation)? Does a user's initial intention change his or her later experience of the site? Does It predict the likelihood that he or she will become an active member or volunteer?
- What accounts for the decline of these sites? What role is played by technological change, ageing or change in the demographics of members? The decision making of the founder? Constraints on human or monetary resources? Something else?
- (Longer term) how can the active user discussions about governance and participation on these sites be captured on Coopedia?

# OCR Progress Report Annex 4

## Collaborative News-Reporting & Citizen Journalism Sites

**Yael Granot**
**April, 2009**

## ABSTRACT

This study examines sites engaged in the collaborative production of citizen journalism or participatory journalism. Each of the cases involve people worldwide volunteering their own written articles, photos or videos in response to current world events. Some of the cases include paid staff writers or aggregated elements from other sources, but all have some distinct component of volunteer news contribution. Comparison across the different sites revealed that motivational appeals to ideals of journalistic excellence were especially widespread, suggesting an area for future research within this domain.

## Domain Definition

The domain of focus here is collaborative news reporting sites, but within an especially narrow boundary. The parameters used eliminate sites like Reddit and Digg, which function merely for the aggregation of popular news stories from various sources. Also excluded from this set were sites specifically focus on *community* news- sites that allows citizens to report on the goings-on in their town or city. The scope for each of these sites was national, or most generally, international.

These sites represent what is known as citizen journalism or participatory journalism. Each of the following cases involve people worldwide volunteering their own written articles, photos or videos in response to current world events. Some of the following cases include paid staff writers or aggregated elements from other sources, but all have some distinct component of volunteer news contribution.

These cases were found in our Phase I master list and through links within Wikipedia's "citizen journalism" article: List of citizen journalism sites (Sourcewatch) and List of participatory news media sites (Open Directory Project). This Wikipedia article defines the domain, citing Bowman and Willis's "We Media: How Audiences are Shaping the Future of News and Information." It explains citizen journalism as:

"non-professionals playing an active role in the process of collecting, reporting, analyzing and disseminating news and information....The intent of this participation is to provide independent, reliable, accurate, wide-ranging and relevant information that a democracy requires."

**Site descriptions**

*Nowpublic* is a participatory news network, claiming to harness the 'wisdom of crowds' by giving people the opportunity to cover the news that is important to them. It allows participants to contribute as editors, writers, bloggers, videographers and on-the-scene reporters. The site's goal is to present "uniquely personal" perspectives on the news. (http://www.nowpublic.com/)

*ePluribusMedia* is a collective endeavor on the part of citizen volunteers to create a space to research and publish stories related to "government propaganda or seeming anomalies disseminated by our news media." It includes four sub-sites: a media page that functions as a weekly news-magazine produced by its users, a community page with forums for discussing citizen journalism and ethics, an "Investigates" page for "vetted" reporters, and a section of timelines related to various research topics (e.g. soldier PTSD reports, Katrina).
 (http://epluribusmedia.net/)

*The Daily Acts* is an attempt at a collaborative news source "produced by its own consumers", focused initially on just the 2008 American presidential campaign. It seems that, though wanting to move beyond this election, by setting its initial parameters too small, the site was destined to failure. This site was deleted some time between the initial coding (early January 2009) and February 1, 2009.
(http://www.thedailyacts.com/index.php)

*Global Reporter* defines itself as a news-related social networking site. Users can post videos related to current issues, friend other users, and share photos and music. While the site identifies itself first and foremost as a social networking site, it encourages the creation of petitions on various worthy causes, though this is a call that seems yet to be answered by users. (http://globalreporter.com/)

*You Scoop It* is an online collaborative news site where users are encouraged to post written articles, videos, and pictures of current events, and to comment on one another's contributions. Still in its alpha testing stage, the site stresses how we all witness newsworthy events every day, and aims to make it easier to share them.
(http://youscoopit.com/)

*WikiNews* is a collaborative global news wiki where volunteers can contribute and edit articles. A Wikimedia Project, it offers its users far more tools that the rest of the cases in this domain. It comes in both print and audio editions, offers Internet Relay Chat for "wikinewsies" to discuss stories, and offers a Newsroom space for users to propose ideas and request collaborators.
(http://en.wikinews.org/wiki/Main_Page)

*Wannabehuman* is a UK-based weblog that promotes citizen journalism. With a centralized screening and editing process, the site claims to be an excellent launching pad for budding journalists to test their mettle. It also claims to offer the direct help of editing

professionals, and sports the motto "virtue is knowledge." Due to the limited, stagnating activity on the site, it is unclear whether this professional support is actually available. (http://wanabehuman.blogspot.com/)

*Orato* is a citizen journalism site that prizes first hand accounts and personal perspectives on current world events. It claims to offer paid commissions for exceptional stories, as well as allows users to tip the authors of exceptional articles using PayPal. Orato attempts to entice potential users through flattery, naming them "genuine experts" and "constant explorers". (http://orato.com/)

*iReport* is a user-generated citizen journalism initiative affiliated with CNN. It aims to connect individuals who "share a passion for the news, to foster unique, independent reporting." iReport stories deemed to be the "newsiest" make their way onto CNN. (http://www.ireport.com/index.jspa)

*Newsvine* is a cooperative news website that invites original work of volunteers as well as syndicated articles from other professional sources. Community is integral to the purposes of the site, and so positive interactions ("Random Acts of Vineness") are valued. The site also claims to offer users a percentage of ad revenue. (http://www.newsvine.com/)

**Survey Analysis**

The results of the survey revealed that the majority of these sites had similar governance structures, with 70% affiliated with an incorporated organization, and of that 70%, 29% of the organizations were not-for-profit. Seventy percent of the cases functioned as distributed newtork governances (core-preiphery type), though they were not the same 7 sites. Additionally, for 90% of the sites tasks were self defined.

In examining the patterns of collaboration, it seems that all of the sites (100%) involved inactive collaboration through collation of independently meaningful modules. This is to be expected, as to make a successful news site one needs to have a cache of articles and/or videos that can stand alone in telling a story. All but one site (90%) involved inactive collaboration through the collation of independently meaningless modules. This is also not surprising, as the majority of these sites allowed for the rating of content or users in one form or another. Finally, only 50% of the sites involved active collaboration between participants. The form this active collaboration took varied extensively, from 'newsroom' pages where users could actively seek out collaborators, to sites that sought volunteer editors and fact checkers, to thriving forums. Despite this collaborative activity, 60% of the sites were coded as having collation of independently meaningful modules as the primary mode of collaboration.

Given that half the sites appealed to functional utility, and 80% appealed to normative incentives or ideals, it is somewhat unsurprising that none of the sites required payment for membership, and all but one allowed for minimal time/effort of contribution. As social ideals were a major motivating tool used, only 1 site resorted to explicit mentions

of "fun/pleasure", only 30% offered individual-oriented social-functional utility, and only 30% made appeals to sociability.

Upon initial coding, I rated 60% successful, 20% no, 20% unclear. While a subjective question, it is reassuring to note that while one site was completely deleted since the initial coding, and this deeper dive allowed for a more extensive look into each site, I would not greatly alter these success ratings. With news sites that aim to follow *current* events, it is especially easy to detect failure when sites have had very few recent postings. Some sites, however, were established too recently for a clear determination of their success. Further, if success hinges on following through with all of the claims they make, it is especially clear after a closer look at each of these sites that the majority of cases in this domain have questionable credibility.

**Closer Observation**

 The high percentage of similarity found in the coding report was generally echoed by closer observation, and stems largely from the narrow defining margins of the chosen domain. At least according to Wikipedia's citizen journalism list there seem to be very few other sites like the ones presented here, that truly try and encourage original citizen reporting on a national/international scope. This set of cases is a distinct and narrow type, which may be why all of the sites were established within four years of each other. Additionally, of the sites who cite a specific geographic location for their organizational offices, nearly half are Canadian. This may speak to cultural imperatives toward certain social ideals or the ways in which a culture relates to its traditional news media.

The sites within this set were all founded fairly recently (2004 and later), and it is understandable that a couple are still in beta testings. One site, YouScoopIt, is even still in its alpha stage. Other than a lack of technological kinks there are other clear differences among more established and more highly visited sites. The more developed citizen news sites stress the importance of tagging content to attract readers. They are more likely to have colloquial guidelines in addition to Terms of Service (Some sites even clearly had fun translating from the legalese: "You acknowledge your sources, and give credit where credit is due. You're a model citizen, sporting your Stetson with pride and honor, pardner."). These well-trafficked sites were more often than not associated with an established corporation, and as such often had a page listing available jobs within that organization. Further, they were able to offer more sophisticated tools to their citizen reporters: while the majority of sites offered RSS feeds, more successful sites were able to offer photo galleries for use, widgets for blog and phone, chat options, browser extensions, and even, in the case of Wikinews, a world news quiz and weather forecast bar. However, more than readership, a site's stated goal had a significant impact on its layout and support.

There was a clear division between sites that claim to be a nobler alternative to traditional media and sites that purport to serve as a crucial supplement- the personal, experiential side of current events. There were also a few sites that fell in between the two, claiming that allowing more people to get involved in the production of the news leads to a more

accurate and balanced result. These self-definitions varied interestingly with revenue sources. In general, a site's revenue sources were rarely anything other than advertising or support of the affiliated corporation. Types of advertising, however, varied tellingly. Sites that most strongly stated ideals of surpassing traditional news media were also the sites whose ad content seemed to more often than not be about other news sources or important global causes. Sites that never claimed such lofty ideals (and, generally, offered monetary incentives) were more likely to invite a wide array of ad content, from weight loss programs to online IQ quizzes.

Overall, sites within this case set up their home pages comparably. They even had similar breakdowns of the sections of their 'newspapers' (e.g. Local News, World, U.S., Sports, Politics, Tech, Entertainment, Science, Business, Health, Odd News). Some claimed to be a great space for budding writers to get experience and start a journalism career through practice and potential networking, but generally, sites that made these claims rarely seemed to follow through on them. WannaBeHuman, for example, states that it pairs people with editing professionals, but proof of these types of connections were unavailable. Additionally, there are sites that try and encourage original content by claiming that only original material will earn top spots (front page), but more often than not, reproduced or linked content can easily get pride-of-place.

**Mechanisms of Cooperation**

Most of the sites in this domain function through distributed network governance, with people at the core establishing certain norms, and all participants contributing to their enforcement. In accordance with this form of government, most sites allowed for self-defined tasks. However, many sites had tips or urgings on the formatting and composition of content. A number of the sites even offered extensive writing guides and advice. While this is clearly in the service of getting people to write original articles, a variety of other types of contributions were possible.

Different Types of Collaborative Activity: For most sites there was some room for both original reporting and "synthesis articles", or summarizing findings from other sources. Additionally, activity on these sites included editing, rating content, posting comments, participating in forums, posting video or image content, sharing articles through other social media platforms, marking other users as your favorites, messaging them, creating a detailed user page, and moderating the content of others. Interestingly, only a small few of the sites allowed for editing or fact-checking of content, which speaks to how much sites like these want to, first and foremost, spur the creation of original, independent content.

Moderation: There was very little variability in methods of punishment or enforcement, with suspension or termination of a user's account as the norm. There were some sites that offered a three-strikes policy, and one that even outlined various categories of flamers, but the majority held to an immediate termination standard. The majority of sites also depended on average users to serve as moderators by flagging aversive content, giving low "trust" ratings, or rating something "fishy." An interesting divergent site was

Newsvine, which specifically gives users the following option: "If you don't want to or aren't able to moderate discussions taking place on your column, uncheck "allow comments." This gave readers the choice of moderating debate proactively by preventing a debate from occurring. Newsvine purposefully puts this option toward the end of its' enforcement section, however, as they see back-and-forth comments on content as a vital element of their site.

Ranking: Most of the sites in this domain allow for ranking, either of content, users, or both.  Ranking of users can vary from a simple aggregation of ratings received on their stories to a more complex algorithm including number of articles contributed, number of comments written and received, and popularity ratings determined by how many fans or followers a user has. Sites like Newsvine make clear how persuasive ranking can be, with an extensive and daunting Leaderboard that tracks everything from ratings by other users to number of comments received to number of links seeded.

Motivational Levers: The table below seeks to capture the various types of motivational levers employed by these sites. Other motivational elements not found in the table include appeals to sense of community, as well as "tangible" praise from other members in the form of online trophies (usually just a graphic that appears on the user's personal page)

## Table of Motivational Levers

| | Year Estab. | Org. location | # of users | Ranking of content | Ranking of Users (not just content) | $$$ Incentives | Appeals to social ideals/ functional utility | Social Networking Features? |
|---|---|---|---|---|---|---|---|---|
| NowPublic | 2005 | Canada | 5,500+ | Can click "recommend" | Yes | No | Yes | Yes |
| ePluribus Media | 2005 | | | Only others' comments | No | No | Yes | No |
| Daily Acts (R.I.P) | 2006 | U.S.A. | | Yes | No | No | | No |
| Global Reporter | 2008 | Canada | ~100 | Yes | No | No | Yes | Yes |
| You Scoop It | 2007 | | | Yes | No | No | No | Yes |
| WikiNews | 2004 | | 83,389 | No | No | No | Yes | No |
| WannaBeHuman | 2005 | U.K. | | No | No | No | Yes | No |
| Orato | 2006 | Canada | 10,000 | Yes | No | Yes | Yes | No |
| iReports | 2006 | U.S.A. | 230,500 + | Yes | Yes | (Wikipedia says Yes) | No | Yes |
| NewsVine | 2006 | U.S.A. | | Can vote to increase rank | Yes | Yes | Yes | Yes |

## Social Ideals as a Motivator

As such a large percentage of the sites employed some explicit appeal to social norms or ideals, this particular motivational tool bears further scrutiny. The very nature of the domain of citizen journalism is the promotion of an ideal, as seen by the definition in the first section, which named the necessity for "independent, reliable, accurate" reporting. To understand the weight these ideals may carry with a potential user, we must explore some examples of appeals these cases make:

"changing the way news is made and distributed… the truth will emerge" (NowPublic)

"Our goal is to shine a light on government propaganda and restore the dignity and integrity of the free press, the lifeblood of democracy in America…to enhance the vitality and effectiveness of the American news media, an institution vital to the future of this nation…to researching issues of common concern and encouraging the highest standards of ethics and journalism" (ePluribusMedia).

"We never want any American to feel as though an agenda-driven media corps has muted his voice ... because every American is the media." (Daily Acts)

"virtue is knowledge" (WannaBeHuman)

"dedicated to aggregating and amplifying voices that supplement traditional media and make it easier to wear someone else's shoes…An accurate, factual and verified but experiential approach gives our audiences the best of both worlds - the macro and the micro point of view." (Orato)

"bring together big and little media in a way which respects established journalism and empowers the individual at the same time."(NewsVine)

It is clear that individuals who register on citizen journalism sites find these messages compelling because, beyond content rankings, they are the main benefit a user receives from his or her contribution. They also relate strongly to another motivator, not found, as such, in the codebook, and not even overtly stated on more than one site within the set. I refer to the ability of these sites to forge a new identity within the user. He or she, without the necessity of formal training, becomes a "reporter," with the responsibility of upholding all of the above ideals.

## Discrepancies & Survey Problems

The majority of discrepancies found between the survey results and information found upon closer observation have to do with the sheer amount of content able to be reviewed. Many things were not apparent within an hour-long coding session that became easily clear with greater scrutiny. Also, there were a number of elements my eye skipped over, as they were not covered in the codebook. For example, I had not previously explored the types of activities allows to registered versus non-registered users. This speaks to a larger problem of our need to determine whether Phase II coders/divers should be required to register for all sites with a dummy email account. I initially concluded that sites within this domain generally told users about all the features they were missing out on by not

registering, but my deeper dive showed that this is not always the case. Further, the amount of material visible to non-registered users must vary widely among different domains.

Other problems with the survey where largely definitional: determining input versus output, applying the "time/location" contingency to the type of reporting done on the sites, and distinguishing between recommendation of content and distributed filtering. Most coding problems seem to relate to technical elements that we have already addressed in Codebook Revision meetings.

One interesting discrepancy, as seen most clearly through The Daily Acts (R.I.P.), is that of time. The time elapsed between the initial coding and the closer observation has a surprising power to change the nature of a site, from its format to tools offered to busineess partnerships formed to its very existence. A more standardized way of dealing with the evolution of a site would be helpful in thinking about how to make such observations.

When one's mission is to click on every possible link within the site, as opposed to answering structured binary questions, one is able to see a wider range of material. Among the things I noticed that did not necessarily come up in the initial coding were more extensive levels of social interaction and praise-giving, the ways in which sections of writing advice/tips can serve as preemptive moderation mechanisms, the types of links to other sites that a given case promotes, and much more. It was certainly easier to see patterns and divergences upon closer observation than through initial coding (though the results from that initial coding certainly helped).

Another important discrepancy to address is not between the closer dive and the coding, but between both of those and the information available on Wikipedia. The setup of our coding tool at times encourages Wikipedia as an additional source of information. And we are clear in the understanding that if there is a conflict between what the site says and what Wikipedia says, we treat the site's claims as fact. However, it would be helpful to think further on what to do when the site make no mention of something and Wikipedia fills in that seemingly unavailable knowledge. For example, iReport was coded as making no appeal to monetary incentives, because no overt mention of such a benefit was found, even on closer examination. The Wikipedia page for iReport, however, listed a number of instances in which users who produced content deemed worthy of CNN were often paid huge sums for the rights to that content. Often generated by individuals who have some interest in a given case, a Wikipedia page can often tell us a great deal about implicit claims made by the site to those in-the-know, information that might not be available to us without weeks and months spent as registered users.

**Questions for Further Study**

- What are the main differences in the design levers at work between sites of this type and sites that involve news aggregation (yet are often still called collaborative news sites)?
- What happens to these mechanisms when we take the scope smaller, to focus on

sites that report events within individual towns and city communities?

- What are the fundamental differences between sites that find it necessary to use monetary rewards and sites that do not? What about sites that offer ranking of users as opposed to sites that only allow for ranking of content?
- How do we determine (and how do users of these sites rate) what is considered 'quality' citizen journalism?
- Must a citizen journalism site have some connection with/influence over traditional news media in order to be considered successful?
- Is there a way to explore whether users that consume citizen journalism either also consume or purposefully avoid traditional news media?
- As almost all the sites use the same, simple enforcement tools, can an even deeper dive determine how effective such mutual flagging is at upholding site standards?

# OCR Progress Report Annex 5

*The Daily Kos*:
**A Case Study of a Large-Scale Collaborative Political Blog**

**Aaron Shaw**
**October 2008**

*Introduction:*

Since its creation, The Daily Kos (http://www.dailykos.com) has evolved into the most heavily trafficked, collaborative political blog on the Internet. Generally characterized as radically left-wing, the site has become the pre-eminent symbol of the so-called "Netroots" movement as well as the broader explosion of online political discourse in the United States. While many online news outlets can match the quantity of traffic on Daily Kos, none have harnessed user-generated content to a comparable extent. The extensive use of distributed information production and filtering on Daily Kos makes it interesting case for political communications research and studies of large-scale networked cooperation.

While a number of research papers and reports in the popular press have analyzed Daily Kos as the standard-bearer of the Netroots, few (if any) have examined the internal functioning of the site community in any detail. As a preliminary step towards more rigorous analysis of cooperation dynamics on *Daily Kos*, this paper attempts an institutionalist sketch of the site, with an emphasis on the structures and practices that facilitate the large-scale information production and filtering that are the community's hallmark. Using a number of qualitative methods, I draw on publicly available information, observation and participation on the site over a 5 month period, as well as a few relevant secondary sources.

The case study is organized into the following sections: (1) basic information and research concerns; (2) overview of the history and evolution of the site; (3) descriptive analysis of its organizational structure; (4) characterization of the technological architecture and software upon which the site is based; (5) examination of the community governance, norms, and culture; and (6) some tentative conclusions and questions for further research.

## I – Basic Information and Research Concerns

*The Daily Kos* has become the benchmark by which other large-scale political discussion sites are measured. As of May, 2008, the site had approximately 168,000 registered users and received between 500,000 and 2 Million hits daily.[45]

More impressive, however, is the scale of collective content production by *Daily Kos* users and contributors. The active community generates approximately 15-20 front-page posts of varying length; over 300 diaries (user-created mini-blogs hosted on the site); and several thousand comments every day. In addition, community members engage in a continuous process of distributed information filtering and discussion moderation. With the exception of front-page posts, all content submitted to the site immediately becomes subject to the site's commenting and recommendation features, through which users contribute feedback on that content's quality. In turn, this feedback is used to facilitate the reading and filtering of information on the site (see Illustration 1).

The site presents a number of research puzzles related to the broader study of networked cooperation and political discourse on the Internet. First, how does the technological and institutional arrangement employed by the site enable distributed production and moderation of information within the community? Furthermore, to what extent do these socio-technical systems enable the site to manage effectively the so-called signal/noise ratio (a fictitious metric of high quality discourse) of contentious political speech?
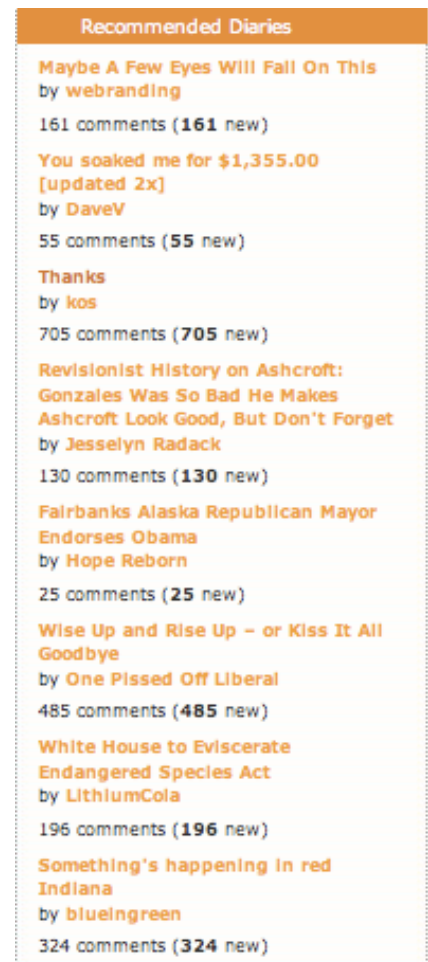


*Illustration 1: Screenshot of "Recommended Diaries" on The Daily Kos landing page.*

---

45      User data from http://www.dailykos.com; Traffic data from aggregate statistics collected by "The Truth Laid Bear" (www.ttlb.com) and Nielsen's "Blogpulse" service (http://www.blogpulse.com/). The mean daily traffic is around 600,000, but it swells during elections and other major political events.

## II – History and Evolution of the site

Markos Moulitsas Zúniga began Daily Kos as a solo-authored weblog on May 26, 2002, advertising the site as a source for "political analysis and other daily rants on the state of the nation." Prior to 2002 Moulitsas had been a prominent contributor to MyDD, a left of center group blog led by Jerome Armstrong. At the time, few collaborative political blogs existed and Moulitsas authored most of the primary content of Daily Kos. Built on the Movable Type blogging platform, the site's early technical architecture and content reflected this hierarchical organization. Occasionally, Kos invited guest bloggers to post his absence or address particular issues, but he remained the sole arbiter of the front page content. Nevertheless, (thanks to Kos' tireless effort and existing readership) the site drew a large audience and many of Moulitsas' posts attracted several thousand hits as well as upwards of fifty comments.

Between 2003 and 2006, the Daily Kos site underwent several major transformations. On October 18, 2003, Kos implemented a technical migration from Movable Type to "Scoop" platform designed for use in the Kuro5hin discussion community. Scoop included a number of reputation and recommendation-based features that the site to incorporate readers as contributors in a more effective, scalable manner. Members of the community could now maintain their own "diaries" (basically, a small scale personal blog) within the Daily Kos site. In addition, Kos formalized a system of "contributing editors" who shared "front page" posting rights and the ability to moderate content. The community's leadership also complemented the Daily Kos site with "dKosopedia," a wiki intended to serve as an informational resource on the American political system and the community itself. While dKosopedia never quite became the encyclopedic repository on of political knowledge that it's creators intended, it remains a definitive resource for the norms and rules that govern the Daily Kos community. This period also marked the appearance of advertisements on the site and the Daily Kos "store," through which fans could buy Kos-labeled clothing and other paraphernalia.

As a result of these technical and organizational changes, the density of both primary (front page) and secondary (diaries and comments) content on the site exploded. This rapid expansion was reinforced during the 2004 election cycle. On the strength of the Howard Dean campaign, the emergent Netroots movement activated its immense organizing capacity and attracted national visibility (Armstrong and Moulitsas 2006). Daily Kos became a symbol of a new kind of politics on the American left. Dean's presidential bid may have fizzled, but it confirmed the status of the Netroots as a new fixture in the American media and political landscape. The waning popularity of the Bush administration following Hurrican Katrina as well as the Democratic Party takeover of congress in 2006 (built on the "Fifty State Strategy" endorsed by the Netroots as well as Howard Dean in his new position as DNC chair) has only reinforced this trend.

Since 2006, the Daily Kos site and community has assumed a relatively stable form. The only substantive shift during that came in 2006 with the creation of Kos Media LLC to

provide a formal legal, financial, and organizational structure for the site and Kos' other business ventures. While responsibility for the site and the bulk of its primary content continues to lie with Kos, CTO Jeremy Bingham and a small team of contributing editors, the majority (by volume) of the writing comes from the community in the form of several thousand comments and diaries contributed daily.

## III – Organizational Structure

Three distinct, but overlapping institutions constitute the "Daily Kos" media ecosystem: (1) Kos Media LLC, (2) the Daily Kos editorial team, and (3) the site's "community." The common factor linking these elements is Kos himself, who maintains a prominent presence on the site while also working as a consultant, activist, author, public speaker, book publisher, parent and recreational composer. Ironically, Kos' personal dynamism and charisma have distracted attention from the fact that the blog - which began as the side project of a single individual - has become a complex organization that has taken on a life of its own.

Kos Media LLC is the privately owned publishing firm founded by Moulitsas in 2006 that is responsible for the technical maintenance, management and oversight of Daily Kos. Publicly available statements on dailykos.com and the dKosopedia indicate that Kos Media currently employs two people: Moulitsas and Bingham (the CTO). The company draws revenue from donations, subscriptions, advertising, and sales of "Daily Kos" branded goods. In addition to providing support for the site and its two full-time employees, Kos Media has also sponsored the "Kos Fellows" program since 2007 in order "to help fund a new generation of progressive activists" (www.dailykos.com/special/about2). In addition, the firm operates two affiliated blogs: Mothertalkers (http://www.mothertalkers.com) and Street Prophets (http://www.streetprophets.com).

As of March, 2008, the Daily Kos editorial team consisted of 21 people including Moulitsas and Bingham (www.dailykos.com/special/about2). The site does not include detailed information on the selection process or hierarchy internal to this group. On a day to day basis, the editors all have front page posting privileges and appear to share other administrative responsibilities related to the site's primary content. The "masthead" indicates that one member of the team acts as an executive editor ("Susan G") and goes on to provide short biographies of each of the contributing editors. Notably, the biographies reveal that all of the editors have college or professional degrees and many maintain careers as lawyers, doctors, scholars, political activists, and writers in addition to their work on Kos. The biographies also mention how long each editor has held their position and whether or not they are "Kos Fellows."

For the purposes of this research, the Daily Kos community encompasses the full spectrum of readers, lurkers, trolls, and contributors that make the site one of the most heavily trafficked and linked-to blogs in the world. Such a diffuse and amorphous group does not operate like a traditional media organization. Nevertheless, social stratifications, technological constraints, and divisions of labor structure the site's production and

consumption in ways consistent with the effects of other organizational forms. Durable and rigid distinctions separate Moulitsas, Bingham, and the editorial team from each other as well as from the rest of the site's users. Additional divisions differentiate the users among four semi-fluid, graduated tiers: (a) front page posters; (b) peer moderators, also known as "trusted users;" (c) registered users; and (d) unregistered visitors. I discuss how the boundaries between these groups interact with the norms, culture, and practices produced on the site in greater detail below. I also analyze the technical platform that plays a critical role in the production and maintenance of the tiers.

As an organization, Daily Kos maintains several semi-formal affiliations which, in conjunction with the blog's internal structure, position it within the American political system and public sphere. Among these, the site's connections to the Democratic party are the most important. While not officially part of the Democrat's national party structure, the blog's leadership and the vast majority of its contributors explicitly identify as Democrats. Daily Kos editors and contributors regularly engage in activities such as campaigning, fund-raising, and volunteering through the site on behalf Democratic politicians. Democratic "special guests" such as political candidates, party leaders, and commentators (e.g. Keith Olbermann) also make semi-regular posts to the site. As a result, it makes sense to think about Daily Kos as part of the universe of organizations affiliated with the Democratic party. Daily Kos's other affiliation stems from its prominence in the "Progressive Blogosphere." Moulitsas, the editors, and many of the site's contributors regularly speak of the site in relation to the "Netroots." While the idea of the Netroots refers to a loosely connected network of American political blogs and not a formal organization, this wider network helps define Daily Kos' public identity as well as its community structure.

**IV – Technological Architecture**

Several technological tools help constitute the Daily Kos community. In addition to the primary platform that runs the site, these tools include the dKosopedia wiki, IRC chat channels, and a series of peripheral email lists. While these programs empower users to perform many roles within the community, the organizational and governance structure of the community as a whole derives primarily from the site's main interface. As a result, I focus my discussion of the site's technical architecture almost exclusively on this interface.

A piece of software known as "Scoop" serves as the platform that runs Daily Kos. As noted above, Daily Kos migrated to Scoop in 2003. However, Scoop's history began several years earlier, in 1999 when a programmer/hacker named Rusty Foster created the project as an open source tool for collaborative discussion moderation. Foster designed Scoop to correct many of the problems he had experienced as a Slashdot member, and he used the new program to create "Kuro5hin" (www.kuro5hin.org), a technology news and discussion community intended as an alternative to Slashdot.

As of 2008, Daily Kos has a community of readers and contributors several orders of magnitude greater than Kuro5in's and is the largest site to use Scoop. The range of

Scoop's capabilities has enabled the site to scale to its current size at the same time as it has facilitated the persistence of certain contradictions and conflicts within the user community. The Scoop website F.A.Q. (presumably authored by Foster) includes the following description of the program:

Scoop is a "collaborative media application." It falls somewhere between a content management system, a web bulletin board system, and a weblog. Scoop is designed to enable your website to become a community. It empowers your visitors to be the producers of the site, contributing news and discussion, and making sure that the signal remains high.[46]

This statement captures the hybrid nature of Scoop. The software's functionality makes it easy for site administrators to distribute responsibility for content production and moderation across the user community, while still retaining some forms of unilateral control. This results in a combination of hierarchical and rhizomatic management practices common to *heterarchical organizations* such as private firms with centralized management that nevertheless employs workers in groups and teamwork (Hedlund 1986).[47]  Analogously, in a *heterarchical* online community, hierarchical relations of moderation coexist with horizontal ones. All users have the power to moderate, but some high-status users have more power than others. This sort of mixed system facilitates the reproduction of institutions, norms, and organizational culture without relying on authoritarian forms of social control. Thus, it also enables the community to scale in a relatively autonomous, self-organizing fashion while still retaining a coherent identity and consistent standards of quality.

Scoop helps Daily Kos decentralize content management responsibilities through a variety of tools or features common to many "social" software applications. Taken together, these features automatically incentivize content contributions that can earn the approval of other users. The system is complex, but not such that it prevents participation. Most fundamentally, the program provides all registered users with the ability to construct a durable and public identity tied to their login name. Once registered and logged-in, users can comment, post diaries (a blog within the site), and customize personal pages. They do this with the knowledge that their work will be visible to any site visitor and (potentially) evaluated by other users. These multiple forms of user-generated content then serve as a foundation for the site's other key social features: recommendation and reputation-building. All registered users have the option to participate in the recommendation system (see Illustration 2, below) . However, recommendations from un-trusted and new users can *only* be positive. New users can only gain trusted status three months after their initial registration with the site.

---

46      http://scoop.kuro5hin.org/guide/C_1_General_installation.html

47      Derek Lakaff (2004) has applied this label to online moderation software and to Scoop in particular. However, Lakaff incorrectly characterizes heterarchy as an absence of hierarchy (2004: 8). He also overlooks the fact that many heterarchical discussion communities rely on strong vertical authority structures for their creation and preservation.

*Illustration 2: Comment display and recommendation options for a registered user. Note the "score" in parentheses following comment titles and "Recommend" check-box at bottom.*

Accumulated over time, recommendations of comments serve as the basis by which users acquire trusted status. In addition, the software automatically "promotes" recent diaries with numerous recommendations on the Daily Kos front page, reinforcing the idea that user-recommended content merits the attention of the community as a whole.

Once users achieve trusted status, they automatically gain access to a wider range of content assessment capabilities. This includes the ability to *un-recommend* content and increased visibility for their new contributions (the result of a weighted average of the reputation established from their recent contributions).[48]

If trusted users do not continue to receive consistent recommendations for their new contributions, however, they lose the "trust" of the site and regress to being "regular" users. So long as they maintain trusted status, however, the system will give them access to extended capabilities. Along with the ability to troll, trusted users can see blocked (or un-recommended) contributions. For a contribution to be blocked automatically by the Scoop system, it must receive at least two "troll" ratings and zero recommendations. Kos also has the power to unilaterally block or censor content they deem especially egregious. Any user whose contributions accumulate a sufficient quantity of "troll" or zero ratings may be automatically banned from the site. Once banned, a user may appeal directly to Kos himself, but the site will no longer allow them to contribute or login with the banned

---

48      See http://www.dailykos.com/special/trusted for more details on the privileges and responsibilities accorded to trusted users.

username.

The use of Scoop and reputational metrics on Daily Kos gives rise to some "reactivity" by influencing the posting behavior of contributors. Indeed, several users have posted tutorials on how to use the site, several of which address strategies for reaching and maintaining trusted user status.[49] While public documents indicate that the software uses an algorithm to calculate user reputations on a periodic basis, the administrators of Kos have "tweaked" the algorithm to reflect their interests. They refuse to disclose either the algorithm or the reputational "rating" of any particular user. As a result, community members possess varying degrees of knowledge about the site's reputational metric, which they refer to informally as "Mojo" – the name of the reputational currency employed by the site prior to the Scoop migration.

## V – Community Governance, Norms, and Culture

Despite the fact that the technological architecture of Scoop elicits consistent patterns of Daily Kos user participation, the software alone does not explain the community's interactions. Among contributors, explicit *formal* and more tacit *informal* norms interact to produce a relatively coherent and durable discursive culture. The formal norms emphasize posting etiquette and standards of quality and honesty over political ideology. In contrast, the informal norms that prevail among users and the site's leadership often include explicit positions on political issues. Contradictions and conflicts sometimes result. When they do, the governance structure of the community becomes more salient as site leaders act to adjudicate conflicts.

*Governance and Formal Norms:*

The practice of governance and formal norm creation on Daily Kos derives from the site's organizational hierarchy. As suggested above, I consider explicitly articulated and enforced norms as "formal." Formal norms are collected in several places: the site's "about" page; an F.A.Q. and other how-to resources maintained on the dKosopedia; occasional blog/diary postings from senior community members and site leaders. These resources generally address issues such as ground-rules for postings, content restrictions, and enforcement practices. In their substance, many of the formal norms derive from Kos' original vision for the site as well as the management techniques he has historically used to manage the site. As a result, the formal norms have sometimes emerged as ad-hoc solutions to a given problem. For example, a diary post from July, 2005 authored by Kos, illustrates how he sought to establish a set of norms around "conspiracy theories":

---

49    See, for example: http://www.dailykos.com/storyonly/2007/4/20/195413/954

**The conspiracists** ⊕

by kos [Subscribe]

Fri Jul 08, 2005 at 11:48:56 AM EDT

Today I did something I've never done before (not even during the Fraudster mess), and wish I'd never had to do.

I made a mass banning of people perpetuating a series of bizarre, off-the-wall, unsupported and frankly embarassing conspiracy theories.

I have a high tolerance level for material I deem appropriate for this site, but one thing I REFUSE to allow is bullshit conspiracy theories. You know the ones -- Bush and Blair conspired to bomb London in order to take the heat off their respective political problems. I can't imagine what fucking world these people live in, but it sure ain't the Reality Based Community.

So I banned these people, and those that have been recommending diaries like it. And I will continue to do so until the purge is complete, and make no mistake -- this is a purge.

This is a reality-based community. Those who wish to live outside it should find a new home. This isn't it.

*Illustration 3: http://www.dailykos.com/story/2005/7/8/114856/8349*

Although Kos later updated to post to clarify that he had reinstated some of the banned users who had appealed to him directly, he made it clear that he would continue to remove all conspiracy theorists and conspiracy theories from the site. At a later date, the original text of this diary post was included in the dKosopedia F.A.Q. section on "Controversial Diary Topics."

Such a trajectory - whereby a norm begins as an action against a particular kind of behavior before achieving formal codification as a written "rule" - may or may not be typical. It is difficult to trace whether some of the formal rules included in the dKosopedia emerged through more collaborative processes of mutual agreement. Nevertheless, the case illustrates how Kos often uses his power and authority on the site to mandate that the community accept his positions on basic standards of discourse and behavior. His ability to ban users and to speak "in the name of the site" empowers him above all other community members. Such tactics may contradict or undermine the collaborative, peer-produced content and editing practices of the site. However, they also serve as a reminder that the site remains an information domain owned and administered by a single individual. When pushed, Kos does not cede ground on this matter.

Ultimately, the sheer magnitude of community-generated content ensures that the enforcement of formal rules is far less than unilateral or uniform. Through my own experience as a participant-observer on the site, I have found that users often deviate from formal norms without suffering any sort of punishment. A whole range of

supposedly forbidden and frowned-upon practices proliferate among the thousands of community members. Examples range from legally risky speech-acts (unsubstantiated ad-hominem attacks on public figures and defamatory statements) to mundane failures to adhere to protocol (ignoring site guidelines on tagging, recommending, and diary formatting).

While these activities technically violate Kos' "diary guidelines," their persistence also reflects his unwillingness to police user contributions in a more thorough way. Kos' statements imply that he believes users' desire to gain the approval of their fellow community members will more effectively limit undesirable behavior than any other enforcement mechanism. Nevertheless, the site software's recommendation and reputation functions allow sub-groups of users to reinforce quirky or questionable practices. In many cases, it would seem that these infractions simply escape the attention of site administrators. The site (and Kos in some ways) still benefit from these "deviant" forms of participation, however, as the revenue from advertisements and hit-based statistical measures of the site's popularity do not account for such nuances.

Thus, even though Kos possesses extreme powers over users and over the creation of formal community norms, the actual "touch" of his authority is generally quite light. Similarly, the dKosopedia may include a number of clearly stated rules governing user behavior, however, the enforcement of these rules remains inconsistent. Such limitations do not mean that the rules or formal norms are "meaningless;" they continue to establish a baseline of acceptability to which Kos and others may, on occasion, refer in their efforts to manage other users' behavior. Yet, the limited ability and willingness of site "leaders" to impose regulations in a hierarchical fashion indicates that informal, internalized norms and cultural values play a much more active role in structuring the behavior of the community.

*Culture and Informal Norms*

In day-to-day interactions between users, contributors, and readers on the Daily Kos site, informal norms (and not codified rules) define the boundaries of acceptable behavior. My experience and research in this area remain somewhat limited because I have not achieved "trusted" user status and therefore cannot see "trolled" or censored comments. Nevertheless, the publicly visible material on the site offers sufficient evidence to support the centrality of informal norms. In general, these norms help establish standards of discourse and stylistic conventions among the site's users. The community also enforces limits on legitimate political speech through informal mechanisms, contributing to the creation of a coherent culture on the site. This culture reproduces itself somewhat independent of the formal rules that "govern" the community in a traditional sense.

As with the site's formal rules, the bulk of the informal norms concern the written conversations conducted among the user community. Areas of activity subject to these informal codes include: the responsible use the site's recommendation and reputation features; the political vision of the community; and the means for resolving internal disagreement. In general, such topics fall outside the scope of the ground-rules

established by Kos himself. As a result, they have become the focus of protracted community *discussion rituals* that serve to produce consensus around certain practices or values. These discussion rituals take several forms: (1) front-page posts; (2) comment threads; (3) diary rescue; (4) so-called "meta" diaries; and (5) internal collective conflicts. The rest of this section considers these different examples of discussion rituals in greater detail.

*Front Page Posts*

Front page posts set the tone for discourse in the Daily Kos community. As both the most visible and accessible posts on the site, the front page establishes the journalistic standards by which the rest of the community members' contributions are measured. While the style and conventions of the Daily Kos front page reflect the general format used by numerous bloggers (reverse chronological order, heavy linking, conversational tone, etc.), several characteristics have come to distinguish Daily Kos posts from the rest of the blogosphere. Among these characteristics, the most central stem from a particularly sarcastic/caustic brand of humor used to mock political opponents and a tendency to support arguments with extensive primary source materials. These traits reflect Moulitsas' personal writing style as well as his direct involvement in selecting writers and texts for the front page. Similarly, the common tendencies of the front page writers structure the site's discursive culture in at least two ways: (1) by attracting new readers who find such writing compelling; and (2) by providing a steady stream of commentary that frames debate throughout the rest of the site.[50] In addition, the fact that Moulitsas selects new front-page contributors from among the ranks of the community at large provides an aspirational incentive for ambitious Kossacks to model their authorial voice on these examples, thereby reinforcing the influence-effects of the front page posts.

*Comment Threads*

Comments represent the largest and most immediate form of user participation on the Daily Kos site. Any post - either on the front page or in the diaries - can receive comments from any registered, un-banned user. Similarly, any comment can receive its own comments in response. A typical front page post receives between 100-200 comments of varying length. On average, diaries receive far fewer comments, however, the top recommended diaries frequently receive over 200 comments. In part, the fact that comments help determine user reputation incentivizes (and perhaps inflates) the large number contributions to the comment threads. Nevertheless, the process of commenting and recommending

While it is not possible to generalize about the specific content of the comments on the basis of my observations, conventions of form and style have clearly emerged. Recurring patterns include satirical jokes at the expense of republicans, photos of pets, antipathy towards the shortcomings of the MSM, and expressions of support for popular

---

50     In this sense, it would be particularly interesting to conduct link analysis and meme-tracking within the Daily Kos site in order to better understand how the community structure relates to the movement of ideas and rhetorical strategies across community members' writings.

congressional democrats. As described elsewhere in this study, some of these conventions stem from formal norms or rules set down by Markos and other senior community members. The repetition of these tropes also derives from the fact that users must secure a measure of popular support (through recommendations) in order to maintain "trusted" status on the site. Neither of these factors completely explains the above-mentioned phenomena however, and many of the conventions reflect continually evolving popular cultures and sub-cultures within the community.

*Serial Features: Diary Rescue and "Cheers and Jeers"*

Markos and the other editors have reserved some of the highly visible front page space for serial "features." Among the serial features, "Diary Rescue" and "Cheers and Jeers" play the most important role in defining and reproducing a distinctive culture on the site. Diary Rescue[51] provides an opportunity for experienced site members and editors to "promote" exceptional diaries to the front page, pointing site visitors to them. While rescued diaries vary greatly in terms of their content, topic, style and author they often model characteristics consistent with those of front page posts. As a result, rescued diaries provide readers with yet another exemplary set of texts from which to learn. The fact that diaries are rescued on a daily basis may provide an added incentive for contributors to attempt to make "rescuable" contributions.[52]

"Cheers and Jeers"[53] is probably best described as a cross between a humor column and a chat-room. By far the most heavily trafficked feature/diary on Daily Kos, "C & J" regularly attracts upwards of five-hundred comments per day. The author, known as "Bill in Portland Maine" (or "BiPM"), publishes a new entry every weekday at approximately 8:50AM EST. Generally, his column includes satirical responses to the previous day's news, miscellaneous asides, and pop-culture references. It follows a regular format and aesthetic that faintly resembles the work of other U.S. political satirists on the left such as Jon Stewart and Steven Colbert. Within 10-15 minutes of BiPM's posting, the daily C & J will have attracted over one hundred new comments.[54] Many of the comments come from regular readers who greet BiPM directly, post their own "cheers and jeers" for the day, or reply to the posts of others. Unlike the rest of Daily Kos, the mood in the C & J comments is expressly not political.[55] Instead, a visitor to C & J is much more likely to

---

51      For a recent example of diary rescue, see: http://www.dailykos.com/story/2008/8/11/232544/133/987/566509

52      I still do not know how the "rescue rangers" who conduct the diary rescues are selected. According to dKosopedia, SusanG, the executive editor, oversees diary rescue, however additional details do not appear to be available.

53      For a recent example of Cheers and Jeers see: http://www.dailykos.com/storyonly/2008/8/12/74425/7737/362/566117

54      While the speed of these commenters is not anomalous by Daily Kos standards, it is nonetheless faster than average front page posts and much faster than other posts on slow news days – in my experience, C&J never has a slow day.

55      One user was troll-rated and reprimanded for discussing the democratic presidential primaries in inflammatory terms on Tuesday, May 13, 2008. See the beginning of the comment thread here: http://www.dailykos.com/story/2008/5/13/85118/7425/333/514287

find pictures of cats (known as "pooties" in the parlance of many of the site users) and children mixed in with friendly jokes. Conversations among sub-groups of users will often carry over from one day's C & J to the next. While political humor is a standby, the sort of contentious debate that defines the other comment threads on the site is notably absent.

*Meta-Diaries*

As a new user of the Daily Kos site, I initially followed the links on dKosopedia to learn how to become an active contributor. The most prominent links in dKosopedia directed me to several "meta diaries" explaining standards for recommendations, politeness, and some of the jargon frequently used by community members. All of the meta diaries included on dKosopedia (also a user-maintained component of the site) have been authored by experienced "Kossacks."[56] Taken together, they represent an active repository of tacit knowledge and opinion freely available to all visitors. The diaries generally consist of focused discussions of technical and social aspects of participation in the Daily Kos community, providing rich information about the community's values, culture, and traditions. Although it is beyond the scope of this overview to go into detail about the meta diaries' content, the process of reading and internalizing their ideas likely serves as an effective rite of passage for many newbies.

*Major Events and Conflicts*

As in many communities, major events and conflicts on Daily Kos provoke reflexive analysis by the site's users. The polemical tone of discourse on the site easily becomes inflammatory in these situations, resulting in heated arguments that threaten the quality of discussion. On such occasions, it falls to Markos and other prominent community members to adjudicate disputes, re-establish standards of conduct, and promote resolutions.

The Democratic Party presidential primaries of 2008 have provided numerous example of these processes in action. During the lead up to the Super Tuesday primaries (February 4, 2008) the polarization between Hillary Clinton and Barack Obama supporters on the site turned ugly. Markos and the site's editors had endorsed Obama months earlier, making no secret of their disdain for some of Clinton's policy positions as well as the strategies pursued by her campaign. These conflicts bled over into the comments and diaries, where Obama and Clinton supporters clashed. Resentful of the opposition of the site's leaders and the widespread pro-Obama sentiment, many Clinton advocates claimed to be the victims of unfair trolling and flaming. On March 14, one of these users proclaimed a "Writer's Strike," urging Hillary supporters and their sympathizers to immediately cease visiting, reading, and contributing to the site.[57]

---

56    The most impressive and extensive collection of meta diaries has been written by "ek hornbeck" and is available at: http://www.dailykos.com/storyonly/2007/2/23/175618/002

57    See: http://www.dailykos.com/story/2008/3/14/20827/4727/132/476843

Responses to the so-called strike were mixed. However, Markos and some of the other senior members of the community took an aggressive stance against the strikers, mocking them and arguing that they had no right to expect consensus on the site, especially given that Clinton opposed some of the central tenets of the community. In the following excerpt from Kos' post in response to the strikers, he emphasizes the significance of Clinton's political failings:

In this quote (and the rest of the post), Kos situates his critique of Clinton and the strikers in reference to the founding ideals of the site, taking the opportunity to re-iterate his vision of the blog. He then applies this vision to the primary elections and shows no sympathy for those who disagree with his perspective.

While a unique occurrence, this particular example demonstrates how Markos can use his position as leader to *affirm and reinforce community standards* in the wake of important events. Markos does not do so alone, however. Substantive posts about the strike received over 1000 comments each, suggesting that the event provoked a profoundly dialogic response among the community as a whole. This proclivity to hash out internal crises and conflicts through on-site conversation demonstrates my broader claim about the role of discussion rituals on Daily Kos.

## VI – Conclusions and Questions for Further Research

This brief overview of Daily Kos leaves many questions unanswered about the nature of the community and its activities. Future inquiries into user demographics, site-visit data, network topography, and user motivations would enhance the skeletal framework I have presented thus far. In addition, insofar as my analysis has derived exclusively from publicly available information on the site, it has excluded the majority of "read only" visitors who come to the site without contributing content. In order to understand the Daily Kos phenomenon more fully - and especially the motivational matrix that drives the site's continuing popularity - it will be critical to assess the experience of this "silent

It is Clinton, with no reasonable chance of victory, who is fomenting civil war in order to overturn the will of the Democratic electorate. As such, as far as I'm concerned, she doesn't deserve "fairness" on this site. All sexist attacks will be dealt with -- those will *never* be acceptable. But otherwise, Clinton has set an inevitably divisive course and must be dealt with appropriately.

To reiterate, **she cannot win without overturning the will of the national Democratic electorate and fomenting civil war, and she doesn't care**.

That's why she has earned my enmity and that of so many others. That's why she is bleeding super delegates. That's why she's even bleeding her own caucus delegates (remember, she *lost* a delegate in Iowa on Saturday). That's why Keith Olbermann finally broke his neutrality. That's why Nancy Pelosi essentially cast her lot with Obama. That's why Democrats outside of the Beltway are hoping for the unifying Obama at the top of the ticket, and not a Clinton so divisive, she is actually working to split *her own party*.

Meanwhile, Clinton and her shrinking band of paranoid holdouts wail and scream about all those evil people who have "turned" on Clinton and are no longer "honest power brokers" or "respectable voices" or whatnot, wearing blinders to reality, talking about silly little "strikes" when in reality, Clinton is planning a far more drastic, destructive and dehabilitating civil war.

*Illustration 4:*
*http://www.dailykos.com/storyonly/2008/3/17/12417/1285/527/4784 98*

majority."

From the perspective of the cooperation research project, a number of uncertainties surround the nature of collaborative activity on the site. Among these, the most central remains the question of what constitutes successful collaboration on Daily Kos? The vast range of contributions, motives, and results prevents straightforward generalization on the basis of my observations alone.

**Sources:**

Hedlund, Gunnar. 1986 "The Hypermodern MNC – A Heterarchy?," *Human Resource Management* 25(1), November, pp 9-35.

Lackaff, Derek. 2004. *Norm Maintenance in Online Communities: A Review of Moderation Regimes*. Excerpt from Unpublished MA Thesis, La Trobe University, School of Communication Arts and Critical Enquiry.

Moulitsas, Markos; and Jerome Armstrong. 2006. *Crashing the Gate: Netroots, Grassroots, and the Rise of People-Powered Politics*. White River Junction, VT: Chelsea Green.