Making Science More Open:
How Crowd Science Can Learn From the Open Source Experience

Chiara Franzoni
DIG, Politecnico di Milano, P. Leonardo da Vinci 32, Milano, 20133,
chiara.franzoni@polimi.it

Henry Sauermann
Georgia Institute of Technology, College of Management
henry.sauermann@mgt.gatech.edu

Abstract

Crowd science is a new way of organizing scientific research and is attracting growing attention from the scientific community as well as science-based firms. Crowd science projects allow contributions from a wide range of participants and openly disclose intermediate knowledge such as data and logs. We develop a framework to compare crowd science to other modes of knowledge production and highlight its key benefits. Drawing on lessons from the Open Source experience, we identify challenges crowd science is facing and discuss organizational and technical solutions to overcome them. With the help of Open Source organizing principles, crowd science promises to deliver significant benefits to society while also offering unique opportunities for firms.

Keywords: crowd science; community-based production; research collaboration; knowledge disclosure; open science; open source software

1 Introduction

An author going by "Players, Foldit" has published one paper in *Nature*, one in the *Proceedings of the National Academy of Sciences* and one in *Nature Biotechnology* – all in just 18 months. These accomplishments are not bad for a freshman, except that this author is not really a freshman. It is a group of people who participate in a game called Foldit.[1] By playing this game, they developed the structure of a protein that can be used for the production of renewable fuel, drugs or specialty chemicals [1].

Previously, they had developed strategies to considerably shorten the path to solving the problem of optimal protein folding [2]. The Foldit game is an interactive 3-D model of real proteins that players can twist and bind with their mouse with the goal of finding the structure that requires the minimum level of energy to prevent the protein from falling apart. There are no specific skills required to begin playing the game because the software automatically computes the levels of energy implied by different configurations and the players quickly learn what protein configurations work. There are currently thousands of players active weekly on the portal.

In January 2009, more or less at the time when Foldit gained the interest of the first gamers, Timothy Gowers, a mathematician and Fields Medalist at the University of Cambridge blogged about a mathematical problem and suggested provocatively to try solving it in a "massively-cooperative" fashion. In several weeks of wiki-type discussion that involved himself, other top mathematicians, students and school teachers, the problem was solved and the work was subsequently published under the pseudonym D.H.J. Polymath [3].[2] Since then, the Polymath project has solved six more problems, always leading to publication under a collective pseudonym.

Foldit and Polymath are two examples of a new way to organize science - what we call "crowd science". In general terms, two features are defining characteristics of crowd science projects. First, crowd science projects are open to a wide range of contributors who often do not know each other in advance and make no formal effort commitments. Moreover, participants may include not only professional scientists but also "citizen scientists". Second, crowd science projects do not only disclose final project outcomes but also openly share intermediate knowledge such as data and logs.

To the surprise of traditional scientists, crowd science is starting to become a viable approach to making major scientific contributions. Top scholars have started to outsource key aspects of their projects to the crowd. Reputable institutions like Oxford University and NASA have become promoters of crowd science projects in several fields of investigation. However, the crowd science phenomenon raises several important questions such as: How does crowd science work? How does it compare to traditional ways of doing science? What challenges does crowd science face and how can we make crowd science more effective? In our attempts to answer these questions, we can draw on many of the lessons provided in the last 15 years by the Open Source Software (OSS) experience, which involves many of the same ingredients.

2 What is crowd science?

In Table 1, we show a small selection of crowd science projects, with no attempt to be complete. The table specifies the primary scientific field of a project, illustrating the breadth of applications across fields such as astronomy, biochemistry, mathematics, or archeology. We also indicate what kinds of tasks

are most common in a particular type of project, e.g., the classification of images and sounds, the collection of observational data, or collective problem solving. This column shows the variety of tasks that can be accomplished in crowd science projects. We will draw on the examples listed in Table 1 throughout our discussion.

Table 1: Examples of Crowd Science Projects

NAME	URL	FIELD	PRIMARY TASK
Ancient Lives	http://ancientlives.org	Archeology	Transcribe
ARGUS	http://argus.survice.com/	Oceanology	Measure & input
Be a Martian	http://beamartian.jpl.nasa.gov/welcome	Astronomy	Identify & flag
Discovery Life	http://www.discoverlife.org/pa/ph/	Biology	Input
Ebird	www.ebird.org	Zoology	Observe & input
Field Expedition-Mongolia	http://exploration.nationalgeographic.com/mongolia	Archeology	Identify & flag
Eterna	http://eterna.cmu.edu/eterna_page.php?page=me_tab	Biochemistry	Game
Foldit	www.fold.it	Biotechnology	Game
Galaxy Zoo	www.galaxyzoo.org	Astronomy	Classify & flag
Great Sunflower Project	www.greatsunflower.org	Biology	Plant, observe & input
Ice Hunters	http://www.icehunters.org	Astronomy	Identify & flag
Moon Zoo	www.moonzoo.org	Astronomy	Identify & flag
Old Weather	http://www.oldweather.org	Climatology	Transcribe
Open Dinosaur Project	http://opendino.wordpress.com/about/	Paleontology	Input
Patientslikeme	http://www.patientslikeme.com/	Medicine	Input
Pigeon Watch	http://www.birds.cornell.edu/pigeonwatch	Ornithology	Input
Phylo	http://phylo.cs.mcgill.ca	Genetics/ bioinformatics	Game
Planet Hunters	http://www.planethunters.org	Astronomy	Classify & flag
Polymath	www.polymathprojects.org	Mathematics	Problem solving
Set quest	http://setiquest.org/	Astronomy	Identify & flag

SOHO Comet Hunting	http://scistarter.com/project/529-SOHO%20Comet%20Hunting	Astronomy	Identify & flag
Solar Stormwatch	http://www.solarstormwatch.com	Astronomy	Identify & flag
Stardust@home	http://stardustathome.ssl.berkeley.edu/	Astronomy	Identify & flag
The Milky Way project	http://www.milkywayproject.org	Astronomy	Identify & flag
Whale Song	http://whale.fm	Zoology	Listen & match

2.1 Crowd science projects are open to contributions from unknown participants

The first distinctive feature of crowd science is that participation of collaborators in the project is open, allowing interested individuals to freely join any specific project and to make the kind of contribution they wish, with the time commitment they are able to make. By participating in a crowd science project, contributors implicitly agree to the unwritten rules of open participation. They agree that an unspecified number of other participants can join. None of those who join have to ask permission and none are asked to grant permission to those who will join after them. Fields Medalist Timothy Gowers' invitation to join Polymath was accepted, among the others, by Terence Tao, a colleague mathematician and Fields Medalist of UCLA. But it also was accepted by a number of other less famous colleagues, as well as schoolteachers and graduate students. Each of them made some contribution, ranging from small comments to major steps towards solving the problem. All were free to jump in the discussion at any point in time, regardless of their skills. Open participation is even more dramatically visible in Foldit, where players can make important contributions even without any background in proteomics. Indeed, approximately two thirds of the leading Foldit players declared to have no more than a high-school chemistry training [2].

Open participation is stretched to the extreme in astronomy. In 2006, NASA launched a project called Stardust@Home. The project sought to identify interstellar dust particles that were collected by the NASA Stardust Spacecraft after its encounter with the comet Wild 2. Candidate particles were as tiny as a micron, which required taking 1.6 million automated scanning microscope images of the entire surface of the collector. To shorten the inspection time, NASA decided to upload the images online and ask volunteers to visually inspect the material and report candidate dust particles. This experiment paved the way for what is now called "citizen science", a particular type of crowd science projects in which the necessary skills are very common and no scientific training is required. As such, citizens science allows laypeople to contribute to science not just by giving money, but also by donating their time and work. Startdust@Home was the precursor of several similar projects in the field of astronomy. Perhaps the most popular one is Galaxy Zoo, a project launched in 2007 by Chris Lintott, an astronomer at Oxford

University.[3] Lintott and his team asked volunteers to visually inspect over 900,000 images of galaxies collected by a robot for the Sloan Digital Sky Survey (SDSS). The volunteers were asked to answer a few simple questions about the morphology of the galaxies shown, such as in which directions the axes of the spirals were rotating. During the first year, Galaxy Zoo attracted almost 150,000 volunteers producing over 50 million visual inspections and classifications [5]. Multiple classifications by different volunteers were used to reduce the incidence of incorrect classifications.

Why do scientists not rely on large and powerful computers to accomplish these tasks? It turns out that in several realms of problem solving, humans are much more effective than computer algorithms. For example, humans have a superior capacity to narrow the space of solutions. This is particularly useful in problems like protein folding, where too many degrees of freedom make complete computation unfeasibly long [2]. Humans also have a superior capacity to handle variability and exceptions, which is particularly useful in classifications of images and sounds. A third advantage of humans is their ability to watch out for the unexpected. This skill is essential in many realms of science. Consider Hanny Van Arkel, a 25-years-old Dutch primary school teacher who became one of the most famous volunteers of Galaxy Zoo when she accidentally came across a bright-blue object near a galaxy she was helping to classify. This object looked different from anything she had seen before, and she quickly reported her finding to the Galaxy Zoo community. Other "Zooites" quickly confirmed her observation and astronomers have since studied this new object – currently believed to be a Quasar Light Echo.[4]

Projects that ask the general public to inspect and classify raw data have since then been used to study galaxies, solar storms, lunar craters, satellites, ice in space, the Mars surface, and the merger of galaxies. Marine biologists are asking volunteers to listen to audio tracks of whale songs and to match similar tracks to gain insight into animals' language. Citizen scientists also help archeologists in their efforts to decipher ancient Greek epigraphs, and climatologists in their efforts to transcribe weather reports from Royal Navy daybooks written during World War I.

2.2 Crowd science projects openly disclose intermediate knowledge

The second distinctive feature of crowd science is that it entails the sharing of intermediate knowledge. This includes raw data such as images, videos, audio-tracks, and arrays but also the data produced by project participants. For example, the Whale Song Project has uploaded a database of audio recordings of whale songs, obtained by combining various independent sources of recordings. Classifications of galaxies performed by the volunteers of Galaxy Zoo were made available to everybody in the summer of 2010 [6]. Projects also share logs, i.e., records of participants' discussions, suggested solutions, and the descriptions of possible strategies and heuristics shared among the participants. The

Polymath project, for example, used several blogs and a wiki to host the discussion about strategies to tackle the mathematical problem.

This high degree of openness with respect to data and logs is very important for the functioning of the project. Like OSS development, crowd science relies on the participation of many and often temporary participants. Given frequent turnover, the codified knowledge recorded in the system as a whole provides the memory to ensure continuity in a project [7,8], enabling new entrants to quickly catch up and become efficient, in spite of their potentially limited time and skills [9]. Making the raw data widely accessible is also necessary because participants are scattered geographically and typically collaborate via web interfaces. The OSS experience suggests that cultural norms, beliefs and values play an important role in supporting this high degree of openness [10,11]. In particular, OSS contributors have recognized the tensions between open sharing that enables all participants to draw upon data and knowledge and limited disclosure, which allows some individuals to appropriate fruits of collective efforts but may hinder the progress of the community as a whole [12]. As a consequence, OSS communities have developed strong norms of reciprocity, where contributors expect free access to work products in exchange for their voluntary contributions [13]. These norms have become increasingly codified using both organizational and legal mechanisms such as open source licenses [14].

Before we delve deeper into the strengths of crowd science, as well as into potential challenges, let us step back and consider how crowd science compares to other modes of organizing scientific activity.

2.3 Putting crowd science in context

Drawing on our earlier discussion of key characteristics of crowd science, we can now classify scientific activity with respect to the degree of openness in project participation and with respect to the disclosure of intermediate knowledge. Figure 1 illustrates this approach, allowing us to distinguish in an abstract way four basic knowledge production regimes.

Figure 1: Taxonomy of Knowledge Production Regimes

		Disclosure of intermediate knowledge (data and logs)	
		Open	Closed
Project Participation	Open	1 Crowd science (& Open Source)	2 Open innovation platforms (e.g., Kaggle, Innocentive)

	Closed	4 Traditional science with disclosure of data and logs (e.g., as required by funding agencies, journals)	3 Traditional "Mertonian" science

Crowd science is located in the first quadrant of Figure 1: projects are open to the contributions of many individuals and intermediate knowledge is openly disclosed. Note that this high degree of openness is also characteristic of Open Source Software development. Given this similarity, our discussion of the promises and challenges of crowd science will draw heavily on the OSS literature. Indeed, we will discuss how many of the organizational features of OSS may support the growth and effectiveness of crowd science. First, however, we will briefly discuss the other cells in Figure 1.

Quadrant 3 captures projects that limit contributions to a relatively small and pre-defined set of individuals and do not openly disclose intermediate knowledge. We call this quadrant "traditional science" because it captures key features of the way science has been done over the last centuries. Of course, traditional science is often called "open" science because final results are openly disclosed in the form of publications [15]. However, while traditional science is indeed "open" in that sense, it is largely closed with respect to the two other important dimensions captured in our framework. This lack of openness follows from the logic of the reward system of the traditional institution of science. As emphasized by sociologist Robert Merton in his classic analysis, traditional science places one key goal above all others: gaining recognition in the community of peers. The primary way to achieve this goal is being the first to present or publish new research results [16,17]. While scientists may also care about other goals, publishing and the resulting peer recognition are critical because more publications translate into job security (tenure), more resources for research, more grants, more students, and so on. Since most of the recognition goes to the person who is "first" in discovering and publishing new knowledge, the institution of science works like a highly competitive system of tournaments [18]. Competition induces scientists to expend great effort on their research work and to disclose the results as quickly as possible. However, competition also discourages scientists from helping competitors, explaining why data, heuristics, and problem solving strategies are often kept secret [19-21]. Granted, even scientists working in traditional science sometimes share intermediate results at conferences or discuss them with trusted colleagues to receive feedback. However, this very limited (and often strategic) disclosure is qualitatively different from the structural openness characteristic of crowd science.

Recognizing that secrecy with respect to data and intermediate results may slow down the progress of science, funding agencies increasingly ask scientists to openly disclose data and logs as a

condition of funding. The National Institutes of Health and the Wellcome Trust, for example, require that data from genetic studies be made openly available. Similarly, more and more journals including the flagship publications *Science* and *Nature* require scientists to publicly deposit data and materials such that any interested scientist can replicate or build upon prior research.[5] As such, an increasing number of projects have moved from quadrant 3 - "traditional science" – into quadrant 4, which could be called "traditional science with mandated disclosure". However, even though projects in this quadrant are more open by disclosing intermediate knowledge after the project is finished, data and logs are still kept secret for the duration of the project. Moreover, participation in a given project remains limited to a relatively small number of individuals, often with pre-existing personal ties.

What about the final quadrant? Projects in quadrant 2 solicit contributions from a wide range of participants. However, intermediate knowledge (and in many cases the final results) is not publicly disclosed. Interesting examples of this organizing mode include open innovation platforms such as Kaggle or Innocentive. Kaggle connects organizations that have a data analysis problem to people who have the required expertise. Customers of Kaggle include firms, government agencies such as NASA, as well as groups of scientists who seek help with problems as diverse as predicting HIV progression or analyzing purchasing data from supermarket shoppers. Inquirers post their problems online as the basis for a competition among problem solvers. Prizes can range from 500 to several millions of dollars. Participants are free to join and offer solutions. The winner will transfer the copyrights of the solution to the inquirer in return for the prize. A similar mechanism for solving a broad range of science and engineering problems has been adopted by the open innovation platform Innocentive. Overall, the projects in quadrant 2 are open to a large pool of potential contributors, but they limit the disclosure of results and data. A main reason for the latter is that project sponsors are often private organizations that seek to gain a competitive advantage by maintaining unique access to research results and new technologies.

3 What are the benefits of crowd science?

In the prior section, we compared crowd science to traditional science as well as to other organizing modes that are less open with respect to project contributions or intermediate knowledge. What are the potential advantages of crowd science's higher degree of openness?

Increasing and leveraging project resources. A first important strength of crowd science projects is that they can mobilize previously inaccessible human resources.[6] Brains are a key input for scientific work, and opening up projects to the contributions of a larger number of participants can dramatically grow the labor force devoted to a problem. This is, after all, why astronomers at Oxford University asked the public for help with the daunting task of classifying thousands of images from the

Sloan Digital Sky Survey. As the Galaxy Zoo example illustrates, crowd science offers the unique opportunity to leverage scarce material resources such as telescopes but also the unique expertise of highly trained scientists by complementing them with a more plentiful resource: the help of thousands of volunteers. Of course, these volunteers will often have less scientific training than traditional scientists and may be individually less productive. However, a large number of small positive contributions can still be enormously helpful, especially if projects are designed such that sub-tasks are efficiently distributed to contributors according to their skill levels. We will return to such organizational challenges at a later point.

Increasing knowledge diversity. Many problem-solving tasks require individuals to draw on prior knowledge and experiences. As such, a diverse set of project participants provides more "raw material" that can be used in the development of new and creative solutions [22-24]. Crowd science can draw on a larger variety of inputs simply due to a larger number of project contributors compared to traditional science projects. More importantly, these contributors often transcend organizational, field, and geographic boundaries. Consider again Polymath, which attracted not only academic mathematicians, but also math teachers, students, and people from many other walks of life, all of whom made important and quite unique contributions to the project.

Recent scholarly work, especially on innovation platforms such as Innocentive, has shown that opening projects to a diverse range of potential contributors may increase performance through a second mechanism: It permits the effective matching between problems and pre-existing solutions [25]. In some cases, problems can be solved by a single person, but the chance that any particular person has the solution is very slim. The larger the number of participants, and the more heterogeneous their knowledge, the more likely it is that at least one of them will have the solution that solves the problem at hand. Thus, by "broadcasting" problems to thousands of individuals, solutions can be found quickly and with little additional effort on the part of the person who supplies the best solution. As a concrete example, consider a problem posted on the Innocentive platform by the Rockefeller Foundation who sought to design a solar-powered wireless router to provide Internet access in developing countries. The winning solution came from Zacary Brown, a Texan software engineer who had already thought about a similar problems in his own effort to design solar powered wireless radio equipment.[7] But even if problems are more complex and "ready" solutions do not exist, openness to diverse contributions can be very useful by allowing the matching between specific project requirements and unique and specialized skills or expertise. A project may, for example, require a rarely used statistical technique. While most traditional teams would not have an expert on board, a crowd science project can signal that need to a broad audience of potential contributors and is likely to find an expert with the required statistical expertise. Moreover, this expert does not have to participate in the whole project but can focus his attention on only

those aspects where his contribution is particularly valuable. He may then move on to another project – perhaps in another part of the world – that also requires this specialized skill.

Better verification. In traditional science, the primary verification of results occurs through review by a small number of professional peers before results are published. While this traditional model has served us reasonably well, it is unable to detect all errors or, worse, cases of fraud and data fabrication [26,27]. Higher degrees of openness in crowd science projects may alleviate some of these problems. In particular, a larger number of individuals who are actively involved in the project are more likely to come across inconsistencies, whether these inconsistencies result from error or mischievous intent. The OSS experience has shown that the effectiveness of a project in detecting, reporting, and even fixing bugs strongly increases with the number of users involved [28]. Moreover, while peer-review does not typically entail a detailed review of the data or lab books, the openness of logs and data in crowd science projects allows others to follow the research process, to trace back the origins of certain claims, and to tie these claims to specific individuals. This transparency ensures that contributors can be recognized for what they have accomplished, but it also ensures that they are held responsible in cases of misconduct [see 29]. Some early examples of debugging in crowd science are visible in the logs of the Polymath project, where the relatively large number of "eyes" ensured that mistakes were quickly detected.

Easier accumulation of knowledge. A final benefit of openness emerges not at the level of a particular project but for the more general progress of science. Scholars of science have shown how the open access to research results allows future scientists to build upon prior work in a cumulative fashion [21,30]. While these discussions typically refer to the disclosure of final project results, additional benefits will accrue if other scientists can build on the intermediate knowledge produced by a project [19,31]. Such benefits are particularly large for data, which are typically costly to collect but can often be used to examine multiple questions. Consider, for example, the Sloan Digital Sky Survey, the Human Genome data, or Census data in the social sciences that have all resulted in many valuable lines of research. But even logs and discussion archives can be enormously helpful for future researchers. Indeed, Polymath members have launched a dedicated website, called Tricki.org, that provides a repository of techniques and heuristics useful for solving mathematical problems at all levels of difficulty. Similarly, Foldit makes available tools that enable gamers to discuss, keep track on progress, and share strategies to tackle problems with other participants.

4 Organizational challenges and solutions

Harnessing the benefits of crowd science requires overcoming several organizational challenges. Fortunately, the OSS experience suggests organizational and technical tools that may help overcome these challenges.

Matching projects and people. One key feature of crowd science projects is their openness to the contributions of a large number of individuals, many of whom may not know each other at the beginning of the project. However, organizational mechanisms are needed to ensure that projects and people find each other and that contributors join those projects where they can be most productive.

Simply speaking, one set of approaches makes it easier for projects to find people. Recall that a key benefit of open collaboration is that a broader base of contributors can add a broader variety of skills, including some specialized expertise that is hard to find but can be extremely useful. As such, a system that collects information on potential contributors and that can be searched for certain skillsets or research interests can be extremely useful. Such a system exists already in Brazil, where the Lattes Platform provides a centralized database collecting research profiles, publications, and other information for virtually all scientists trained or working in that country.[8]

A second approach makes it easier for scientists to find projects by aggregating and disseminating information on ongoing or planned projects. This approach has been used successfully in OSS development where websites such as Freshmeat.com, Koders, or Lwn.net provide space for posting news and announcements, forum discussions, as well as indexes and search tools. Programmers get to hear about an OSS project because they read news about its launch, or because they post inquiries and receive answers, or because they search in old posts and read past discussions. In the crowd science space, websites such as scistarter.com now offer searchable databases of a wide range of projects, allowing interested individuals to find projects that fit their particular interests and level of expertise. Project information can also be exchanged more informally on social networks such as researchgate.net or in forum-like discussions between the readers of scientific articles at open access journals such as PLoS.

The best way to solve the matching problem, however, will be to bring projects and people together on larger project hosting platforms that utilize a common infrastructure while also developing a distinct brand recognition. In OSS development, such platforms include sourceforge.com, Savannah and BerliOS. In return for web traffic, these websites provide free space to quickly set up a project, storage capacity to share and backup files, and tools to facilitate and record discussions. They also provide statistics of activities that are extremely useful for a potential contributor to understand the progress of a project. Once such a platform has reached a critical mass, it can become a large repository of code, patches, instructions and data that programmers browse in search of pieces useful for their work [9,12]. More users, in turn, produce new content, leading to further growth. Similar project platforms are emerging in the crowd science realm, though still at an infant stage. By way of example, the Galaxy Zoo project has evolved into the platform Zooniverse, which now hosts over ten projects in the areas of space, climate, humanities, and nature.

Modularization. Common wisdom has known for a long time that "too many cooks spoil the broth". This saying speaks to the fundamental challenge to coordinate the contributions of multiple individuals, and these challenges grow exponentially as the number of contributors increases. How can crowd science – which seeks to exploit scale and skill diversity and a globally distributed participant base – overcome these challenges? Open source software provides part of the answer by illustrating the power of modularity. The basic idea of modularity is that a large problem can be divided into many smaller problems, plus a strategy (the architecture) that specifies how the modules fit together. The OSS intuition is that modules can be designed to have minimum interdependencies with one another, allowing for a greater division of labor and parallel work. In addition to allowing for faster development, modularization also allows learners to try their skills without risking to undermine key functionalities [7,32] and permits greater customization to a wide pool of needs through module recombination [33]. We can see modularity in many crowd science projects as well. For example, Galaxy Zoo keeps the design of the overall project centralized and provides less-skilled contributors with a small piece of the problem ("inspect this image and determine the shape of the galaxy") so that contributors can work independently and at their own pace. Of course, not all problems can be easily modularized, and projects with highly interdependent components are unlikely to be able to use as many participants as projects with a higher degree of modularization.

Integration of contributions. Just as important as distributing tasks to different contributors is the effective integration of the resulting contributions to answer the original research question. As the number of contributors grows, the challenge is to keep track of the various contributions, to pick the creative ones, and to combine them in a way that ultimately addresses the problem. In highly modularized data collection tasks, such as the Zooniverse projects, individual contributions can quite easily be integrated into larger data sets. Similarly, in some problem solving tasks, each contribution is a stand-alone solution and standard evaluation criteria can be used by the project leaders or by the community to judge the merits of the contributions. In the early days of Linux, for example, programmers submitted new modules to Linus Torvalds who would incorporate deserving modules into the source code. In open innovation platforms such as Innocentive or Topcoder, scoring algorithms use objective (though sometimes secret) criteria to rank solutions submitted by the various contestants.

The biggest challenge is the integration of contributions in complex problem solving tasks such as Polymath, where the contributors seek to develop a single solution in an interactive and collaborative fashion, e.g., through an ongoing discussion. In such cases, much of the integration is done informally as participants read each other's contributions. However, as the amount of information that needs to be processed increases, informal mechanisms may miss important contributions while also imposing large costs on project participants. After some time, these information processing costs may deter new entry

[4]. While filtering and sorting mechanisms may lower these costs to some extent, we expect that highly interactive problem solving will best be performed in relatively small groups (compare Polymath's two dozen contributors with Galaxy Zoo's 200,000).[9]

Project leadership. In some cases, such as Polymath, it may be sufficient to post the problem and let the crowd develop the solution in a more or less self-organizing fashion. However, complete decentralization is likely to be problematic for larger projects, for more complex problems, when contributors do not have the specialist knowledge to know what needs to be done, or when there is disagreement about how to proceed. In OSS development, for example, unresolved conflicts about how projects should proceed can lead to "forking", i.e., the development of competing project paths that can lead to a loss of focus and incompatible solutions [34]. Thus, some degree of project leadership is typically needed for the growth and effectiveness of OSS communities [35]. Such leadership does not have to involve a large number of people. Complicated projects such as Linux have shown that architecture and kernel design can be performed by a relatively small and tightly knit group of top-notch programmers, while a large number of people at all levels of skills can carry on and execute smaller and well-defined modules [14]. Emerging crowd science projects such as Galaxy Zoo or Foldit show a similar pattern: These projects are led by well-trained scientists who formulate important research questions and design methodologically sound experiments. Foldit also illustrates that this kind of leadership is not always exercised verbally but can instead be incorporated into technical infrastructure. More specifically, the Foldit software embeds important "rules of the game" right into the software, ensuring that participants perform only operations that project leaders have determined to be consistent with the applicable laws of nature.

Where can leaders come from? While traditional science assigns leadership positions primarily to highly accomplished scientists in formal positions of authority (i.e., lab heads, chaired professors), crowd science can see different types of leaders. They may be professional scientists who provide the scientific and technical skills that enable others to participate, as in the example of Galaxy Zoo. They may also be designers of collaboration tools, who may have less of an interest in a particular content domain per se, but who have – often through experience – built expertise in crowd science project management (i.e., some of the employees at Topcoder or Innocentive fulfill that function). And of course, leaders may emerge from the larger "crowd" as a project grows. Indeed, the OSS experience has shown that leadership should be thought of as a dynamic concept. In particular, technical leadership is generally prevailing at the infancy of a project, but other skills such as communication, participative style and capability to organize become credited bases of authority as the project grows and matures [35,36].

5 Motivation and Incentives

5.1 Why would people participate in crowd science projects?

OSS research has highlighted a variety of different motives and incentives underlying programmers' willingness to invest time and effort in software projects [37]. In the following, we leverage that research to discuss different types of motives and incentives that may drive crowd science participation now and in the future.

Challenge, play and social interaction. A first type of motivation highlighted by OSS research is intrinsic motivation. Intrinsically motivated people participate in a project because they enjoy the intellectual challenge of a task or because they find it fun [38-40]. Similarly, the feeling of accomplishment and competence created by successful task performance may be a powerful intrinsic motivator [11,41]. Intrinsic motives are likely to be salient to many of the crowd science participants as well: playing Foldit and Phylo projects is simply fun, and achieving an excellent score provides a feeling of pride and satisfaction.

While programmers (and scientists) are not known for their social skills, they may also enjoy the social dimension of collaborative projects [41]. In particular, scientists may feel good about being part of a particular project community and may enjoy personal interactions with other participants. Indeed, the conversations captured in project logs such as Foldit, where many participants teamed up to exchange strategies and compete in groups, reflecting not just enjoyment from gaming, but also a collegial and "social" element.[10] Of course, the social element of play has proven enormously powerful in the realm of commercial games, where companies such as Sony, Microsoft, and Zynga have created large networks of players who regularly interact over the Internet.

Knowledge and understanding of particular topics. While challenge and play can be experienced in several different contexts, many crowd science participants have strong interests in particular topics. For example, many amateur astronomers contribute to citizen science projects in astronomy (Galaxy Zoo, Moon Zoo, etc.) because they like to see pictures of the sky, the moon and other space objects. Interest in a particular topic and the desire to learn more about nature have driven scientists for millennia [42]. While this motive is "intrinsic" in the sense that it is not dependent on any external rewards, it is distinct from challenge and play motives in that it is specific to a particular topic, potentially limiting the scope of projects a person will be willing to participate in. To illustrate, someone who likes challenge and play may contribute to projects in biology and to projects in astronomy, as long as these projects provide a playful interface or pose challenging puzzles. In contrast, someone with a special interest in astronomy may not be willing to contribute to a biology project. Recognizing the importance of this motive, some platforms such as Galaxy Zoo have enriched the work by providing scientific

background information and interesting trivia to project participants. An especially powerful version of interest in a particular problem is evident in the growing number of projects devoted to the understanding of particular diseases. Many of these projects involve patients and their relatives who have a very strong personal stake in the success of the project, leading them to make significant time commitments.[11]

Learning to be a scientist. OSS research has shown that some less experienced contributors participate because they learn from doing real programming, from answering the requests of help [39,41] or from the feedback they receive from experienced programmers [28,43,44]. Similar motives may be at work in collective problem solving projects such as Polymath: by following the comments and explanations posted in the blog, graduate students in mathematics could see live how top-notch mathematicians approach problems, try and abandon different solutions, prioritize certain areas of the problem before others, proceed, stop and go backwards. The project provided students not only with a proved theorem, but also with a lesson on problem solving. This makes participation in a crowd science project such as Polymath an exceptionally powerful learning experience, potentially attracting a large pool of students and practitioners. Of course, learning also benefits the project itself by increasing the productivity of project participants.

Recognition and career concerns. Many programmers participate in OSS development because their name will show up in the list of software contributors and this will help their career, for example when seeking a new job [38,45,46]. Career-related motives are unlikely to be important to citizen scientists whose primary career does not depend on scientific contributions. As discussed earlier, however, the motive to earn peer recognition has always been central in the institution of "traditional" science and may be a powerful motivator for professional scientists to start crowd science projects. The Polymath or similar projects can also be a unique occasion for students and other participants to get in touch with established scholars in their community and make themselves noted. Participants can also list contributions in their CV, potentially sending valuable signals to potential employers. Interestingly, some crowd science projects such as Galaxy Zoo and Foldit have resulted in authorship for individual contributors, but others instead name the group as a whole such as in the case of "D. H. J. Polymath". While the use of group pseudonyms is consistent with the spirit of a "communal effort", it may reduce project participation by professional scientists who need publications to succeed in the traditional institution of science.

Use and re-use. A particularly important motive in the OSS context is the need to use the software one helps create [38,43]. Similarly, many programmers re-use pieces of code developed by other contributors to build programs that fit their own customized needs [9,47,48]. Use and re-use may also be important motives in the context of crowd science. We discussed earlier that crowd science entails the free sharing of intermediate knowledge and of final results with the project community and the broader

public. Each of these outputs may be useful for scientists working on related projects, and involvement in the source project can provide faster access as well as a deeper understanding of the nature and usefulness of the material. For example, a proved theorem can be used in solving other mathematical problems. Similarly, data created in a particular project may provide the basis for a variety of follow on studies by project participants.

Money. Finally, many programmers participate in OSS projects because they are explicitly paid to do so: they are employees or consultants of companies [41,43]. While paid participation in crowd science projects is currently the exception, we expect it to become more important as firms find ways to profit from an involvement in crowd science.[12] Indeed, firms did not play a major role in the beginning years of OSS, but have developed several OSS-based business models over the last decades [49]. Why would firms want to get involved in crowd science? What could their involvement look like?

5.2 Firm involvement in crowd science

While many firms use open innovation platforms such as Innocentive (located in quadrant 2 of Figure 1), company involvement in crowd science is currently much more limited. This lack of participation may reflect the difficulty of appropriating financial returns if data, logs, and research results are disclosed openly. We expect, however, that firms will find viable ways to get involved in crowd science. After all, companies have over the years become important contributors to Open Source Software Development [50]. Statistics show, for example, that about 75% of the development of the Linux kernel is done by paid employees of companies [49].

Commercial upgrades and proprietary valued-added. Some OSS companies such as DotNetNuke make money by offering a free version of an open source project to the broader community while offering an upgraded version or proprietary add-ins to paying customers. Similar business models are emerging in crowd science. Consider, for example, the for-profit company Patientslikeme, which invites patients to share their health data and medication regimes on an online platform. All members of the platform can see each other's information, and the website publishes basic statistical analyses such as trends for particular diseases or aggregate information on the subjective effectiveness of certain drugs. Patientslikeme employs several professional scientists who design the data collection and write academic papers using the member data [51]. In addition, Patientslikeme invites external researchers to propose projects or to use existing patient data for academic studies. At the same time, however, the company sells detailed patient data to customers such as pharmaceutical companies who may use the data to speed up drug development or to track adverse effects. Even though Patientslikeme is not completely open (i.e., general users can see only "data light" but not the raw data available to paying customers or to academic

researchers), it demonstrates how a hybrid model can provide significant degrees of openness while also allowing firms to appropriate value.

Providing platforms and infrastructure. Companies such as sourceforge.com operate successfully in the OSS realm by providing free services such as troubleshooting information, project hosting, and file storage in exchange for web traffic. Similar websites are emerging in the crowd science world. For example, researchgate.com and lablife.com provide basic functionalities helping scientists to set up a project homepage, share information, receive and post feedback, advertise their research and show their expertise to people in search of collaborations. Addgene is a Cambridge-based company that provides important infrastructure services by storing and verifying biological materials and by distributing these materials (for a fee) to the broader research community.

Our earlier discussion suggests that crowd science platforms will play an increasingly important role in matching people and projects and in providing infrastructure. While such platforms may be developed by crowd science participants themselves (e.g., the Zooniverse platform), they may also provide opportunities for firms. After all, while crowd science projects may be committed to open participation and the sharing of results, this doctrine of openness does not automatically extend to all of the infrastructure and research technologies. Indeed, commercial firms have for a long time sold valuable research services and infrastructure to the scientific community. Of course, for-profit platforms have to provide superior value if they intend to compete successfully with free alternatives. A particular appeal of platforms is that they are likely to benefit from economies of scale and network effects, suggesting that early movers may gain significant competitive advantages.

Re-use and complementary products. Firms may also get involved in crowd science to use data and research results in their own products and services. Consider Survice Engineering, a company specialized primarily in military engineering solutions, including hydrographic surveying products and services. In 2009, this company started ARGUS, the Autonomous Remote Global Underwater Surveillance System. This crowd science project asks participants to collect data on seabed depth in costal areas by installing a sampling device onboard of their vessels. This device transmits information about seabed depth and the ship's position as a vessel navigates. In addition to drawing on a broad base of participants, the ARGUS project also openly shares the resulting data – any mariner can download the data and a viewing tool on the ARGUS website. However, the firm is also uniquely positioned to profit financially from the project, e.g., by using the data in the development of commercial products such as measurement systems or 3-D modeling services, or by selling the data as part of more sophisticated navigation systems. This example illustrates that even data and research results that are made available "for free" can allow a company to derive financial returns if the firm controls unique complementary resources and capabilities that give it an advantage over competitors. Of course, by taking an active role

in the crowd science project, the firm may gain early access to the data, may shape the project path, and may gain a better understanding of the project results. Firms such as IBM and Intel have for many years realized similar benefits from their involvement in OSS Development [52].

5.3 (How) can different motivations coexist?

Most crowd science projects will not be able to succeed based on just one of the motivations discussed earlier; rather, they have to involve contributors with a diverse set of motivations. One reason for this claim is our earlier observation that a key advantage of crowd science projects derives from diversity in knowledge, backgrounds, and skills. Such diversity is likely to be related to different reasons for participation as well. For example, professional scientists may contribute to gain recognition and career benefits, while citizen scientists are more likely to participate for fun, enjoyment, or the desire to learn about a certain knowledge domain.

A second reason is that projects may need the support of different types of individuals over their life cycle. Historical studies of OSS projects have shown that many projects require a critical mass of contributors to create awareness and attention from other potential joiners. Until a critical mass is reached, the risk of wasting time in a project headed for failure is very large. Hobbyists who enjoy the process or students wanting to maximize learning are more likely to take that risk than professionals who need results [41], while professionals tend to become more active as projects grow and stabilize [14]. While some crowd science projects may have a similar pattern of involvement (i.e., those started by patient groups), other projects such as Galaxy Zoo were started by professionals but could only grow by attracting citizen scientists who were driven by very different motives. Either way, crowd science projects face the challenge that contributors may have very different motives and potentially conflicting goals.

A major contribution of OSS research is to show that different motivations can co-exist, and that potential incentive conflicts can be mitigated using contractual mechanisms. In particular, it is important to recognize that there are several different rights involved in the production of knowledge. These rights include, for example, the right to be regarded as the author, the right of using a product that was collectively produced, the right to make derived works, and the right to impose further obligations on work derived from project results (e.g. demand that derived work be publicly distributed) [53]. OSS research has shown that unbundling these rights can help in attracting diverse groups of individuals because at least some of these rights are not strictly rival and can be enjoyed at the same time [12].

To illustrate, let us consider the example of Solar Stormwatch. This project asks people to watch videos of solar storms and to tag characteristics such as the inception point or the maximum reach of the storm. The small group of lead scientists on this project includes professional scientists working at a government lab as well as a PhD student at Imperial College, London.[13] Let's assume that these scientists

are primarily motivated by the desire to write scientific papers based on the data resulting from the project. Suppose now that a company that engineers solar panels for satellites would want to use these data to enable its equipment to detect the inception of a solar storm. The company may be willing to participate in the project to speed up the completion and release of the dataset, e.g., by paying an employee to work on the project. However, it will do so only if it is ensured access to the resulting data and if it can incorporate data into its computer algorithms and keep the algorithms private. There is little conflict between company's plans and scientists' desire to publish the data or papers based on the data. Now consider a third party, namely another team of astronomers who need data on solar activity for their own research. These researchers will be willing to help with the project if the open and timely release of the dataset is ensured. If these researchers are working on a similar problem as the Solar Stormwatch lead scientists, the two teams are directly competing. The prospect that competing scientists may join and gain early access to data could thus be a potential detriment to scientists' starting a crowd science project in the first place. However, both teams of scientists should be willing to participate if they expect to pursue non-competing research questions using the data.[14] Finally, consider a fourth set of contributors – citizen scientists who simply enjoy watching videos of solar storms and learning more about this exciting phenomenon. In principle, this latter group of contributors should not care who gets credit for scientific results from the project, and they should also not be opposed to a company creating useful products based on the resulting knowledge. However, these participants may nevertheless feel "exploited" if they learn that others derive personal gain from their efforts.

Taking inspiration from existing OSS license arrangements, contractual mechanisms can be designed to incentivize all parties in our example to participate while mitigating potential goal conflicts. For example, the founding team could reserve the right to use the data for particular pre-defined research questions, ensuring that the lead scientists have incentives to invest the time and resources required to run the project. At the same time, the lead scientists would commit to disclose the data openly for other uses – providing incentives for the second team of professional scientists. The license could also specify that algorithms derived from the data may be incorporated in commercial products, incentivizing the firm to participate in the project, without reducing the incentives for either group of scientists. Citizen scientists primarily need the permission to use the project infrastructure – which the organizers should willingly provide in return for free help.[15] In order to minimize the risk that citizen scientists feel "exploited", professional scientists should commit to recognize the important contributions of other project participants, e.g., in the acknowledgement sections of journal articles. Similarly, firms seeking to benefit from project outputs can commit to contribute back to the community, e.g., by paying for project infrastructure that makes it more "fun" for those who volunteer their help.

Of course, such licensing schemes may face various challenges. Among others, it is not clear how well license contracts can be enforced, although some crowd science projects such as Zooniverse already require contributors to login and to accept certain terms and conditions. Moreover, there is a danger that rights become too many and too narrow and hinder the emergence of unexpected research trajectories [see 54]. Most importantly, there may be a resistance in the science community to using licensing agreements since such agreements may seem to contradict the idea of an "open" science. However, it is important to recognize that well-designed agreements would enable the participation of a broader set of participants, the creation of data that would not normally exist, and the sharing of data that would otherwise be kept secret by scientists concerned with winning the race for priority.[16] Based on the OSS experience, the scientific community may find it useful to develop a number of standard licenses or a code of conduct for the design of new crowd science licenses.[17] National granting agencies or scientific associations such as the National Academies could also play an important role in developing such tools.

Table 2 suggests organizational levers that address the various types of motivations and can mitigate goal conflicts among project participants.

Table 2: Organizational Levers to Increase Project Participation

Types of motivations	Levers to improve participation
Challenge, play, and social interactions	Provide feedback on performance (absolute or relative to other "competitors"); Provide entertainment; Minimize entry cost; Design mechanisms for human interaction (blogs, mailing lists).
Knowledge and understanding	Integrate external content on the topic; Provide feedback on project results; Ensure that results have impact on practice (e.g., patient welfare).
Learning to be a scientist	Ensure easy access to logs and problem discussions; Add content on research methods and design; Get senior scientists involved; Provide "safe zones" for learning and experimentation (e.g., via modularity and redundancies).
Reputation and career	List contributors by name; Ensure that individual names appear on publications; Design licenses that limit competition between follow-on research projects.
Use and re-use	Commit to fast and open release of data and results; Design license that limit competition between follow-on research projects; Design modules to facilitate re-use.
Money	Pay for time or pay for specific contributions - especially useful for "boring"

	contributions that would not otherwise get done; Use licenses that allow for commercial use of results.

6 Making sure it is still "Science"

Open Source and crowd science share many similarities. However, there is an important difference regarding how results – computer code in the former and scientific results in the latter – can be verified and evaluated. In OSS, the primary goal is to develop software that fulfills certain functions and that can be incorporated into follow-on work. As such, one important quality check is whether the code "works" or whether there are bugs that need fixing. Future programmers are unlikely to incorporate code that does not work. Things are more complex in science, where the "truth" of scientific findings does not reveal itself as easily [see 21]. Since it is often not easy to tell directly if scientific results are accurate, the scientific community places great emphasis on the way in which research results are generated, and the traditional peer review process focuses on the question whether a result was obtained using appropriate methods [55,56].

The important role of the process by which results are obtained explains some critics' concern that crowd science projects do not always follow the scientific method. This concern is particularly relevant for citizen science projects whose contributors are not trained scientists. Some citizen science projects led by patients, for example, do not use the experimental designs typical of traditional studies in the medical sciences, making it difficult to interpret the results.[18]

To ensure that crowd science meets the rigorous standards of science, it will typically be necessary that trained scientists are involved in the design of experiments or otherwise ensure that the project follows the scientific method. Consider again Galaxy Zoo, where a relatively small number of professional scientists decides how data will be collected and analyzed, but where a large number of citizen scientists help with the actual data collection. Few of the citizen contributors will be familiar with the methodological details – but the overall project structure sensibly allocates that responsibility to the professional scientists. It may also be possible for crowd science platforms to provide "scientific consultants" who advise (and potentially certify) citizen science projects. To some extent, rigor and standardized scientific processes may even be embedded in the software and platform tools that support a crowd science project.

We argued earlier that high degrees of openness throughout a crowd science project and the involvement of professional scientists are likely to improve the verification of results. It is nevertheless important that final research results are also evaluated independently before they are officially included in the body of scientific knowledge. In traditional science, peers perform this important task by reviewing other scientists' work before results are published in journals. To the extent that crowd science results are

published in traditional journals, this review process still applies. However, results are increasingly disclosed through non-traditional channels such as blogs and project websites. The question whether and how such disclosures should be verified and certified is an important area for future scholarly work and policy discussions.

As this article shows, crowd science represents a unique opportunity to advance the progress of science for the benefit of traditional scientists, citizen scientists, firms, and the broader society. While this nascent approach faces numerous challenges, the OSS experience provides useful insights that can be leveraged to understand and improve the effectiveness of crowd science projects. And of course, we should watch carefully as scientists of all types continue to experiment with new and creative ways of organizing scientific discovery.

ENDNOTES

[1] For an overview on the project, visit the project website: http://www.fold.it/portal/
[2] D.H.J. stands for Density Hales-Jewett, a part of the problem they solved. The history of Polymath is detailed at http://michaelnielsen.org/polymath1/index.php?title=Main_Page
[3] A detailed account of both the Polymath and the Galaxy Zoo Project is given in Nielsen [4].
[4] http://www.galaxyzoo.org/story
[5] http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml#dataavail; http://www.nature.com/authors/policies/availability.html
[6] We have seen this potential play out in OSS Development. For example, once Linus Torvalds had posted his code in 1991, increasing numbers of programmers joined the effort, resulting in a tremendous growth of the Linux system, reaching a scale that is difficult to match using other organizational paradigms. Moreover, much of the work has been supplied for free by volunteer contributors – contributions which some calculations value at over 3 billion dollars. http://linuxcost.blogspot.com/2011/03/cost-of-linux.html
[7] http://blog.innocentive.com/2009/02/04/im-a-solver-zacary-brown/
[8] http://lattes.cnpq.br/
[9] http://michaelnielsen.org/blog/the-polymath-project-scope-of-participation/
[10] http://fold.it/portal/blog
[11] http://online.wsj.com/article/SB10001424052970204621904577014330551132036.html
[12] Of course, we can think of professional scientists who participate in CS projects to advance their careers as also motivated by money, though more indirectly.
[13] http://www.solarstormwatch.com/mission_briefing
[14] This discussion suggests that crowd science projects may be more viable in "general" research areas that allow the re-use of data for several non-competing research streams.
[15] It may seem trivial that project leaders will allow free access to the platform in order to increase the "free" labor at their disposal. However, projects may be exciting enough that citizen scientists are willing to pay to help. After all, people pay for other types of online games, people pay for access to scientific results (e.g., National Geographic Subscriptions), and people are willing to incur financial costs for volunteering in other settings (such as mission trips). Indeed, a Puget Sound beach watcher crowd science project charges a moderate participation fee (see scistarter.com).
[16] Moreover, informal agreements about the assignment of credit and use of jointly produced data are already common in traditional scientific teams.
[17] See for example the repository of the Open Source Initiative, which has elaborated ten requirements for software to be considered open source, http://www.opensource.org/osd.html
[18] http://online.wsj.com/article/SB10001424052970204621904577014330551132036.html

REFERENCES

1. Eiben CB, Siegel JB, Bale JB, Cooper S, Khatib F, et al. (2012) Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. Nature Biotechnology 30: 190-192.
2. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, et al. (2010) Predicting protein structures with a multiplayer online game. Nature 466: 756-760.
3. Polymath D (2010) Density hales-jewett and moser numbers. An Irregular Mind 21: 689-753.
4. Nielsen M (2011) Reinventing Discovery: The New Era of Networked Science: Princeton University Press.
5. Lintott CJ, Schawinski K, Slosar A, Land K, Bamford S, et al. (2008) Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. Monthly Notices of the Royal Astronomical Society 389: 1179-1189.

6. Lintott C, Schawinski K, Bamford S, Slosar A, Land K, et al. (2010) Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies. Monthly Notices of the Royal Astronomical Society.
7. Von Krogh G, Spaeth S, Lakhani KR (2003) Community, joining, and specialization in open source software innovation: a case study. Research Policy 32: 1217-1241.
8. Frakes WB, Isoda S (1994) Success factors of systematic reuse. Software, IEEE 11: 14-19.
9. Haefliger S, Von Krogh G, Spaeth S (2008) Code reuse in open source software. Management Science 54: 180-193.
10. Stewart KJ, Gosain S (2006) The impact of ideology on effectiveness in open source software development teams. Management Information Systems Quarterly 30: 291.
11. Osterloh M, Rota S (2007) Open source software development--Just another case of collective invention? Research Policy 36: 157-171.
12. Bonaccorsi A, Rossi C (2003) Why Open Source software can succeed. Research Policy 32: 1243-1258.
13. O'Mahony S (2003) Guarding the commons: how community managed software projects protect their work. Research Policy 32: 1179-1198.
14. Shah SK (2006) Motivation, governance, and the viability of hybrid forms in open source software development. Management Science 52: 1000-1014.
15. David P (2008) The historical origins of "open science". Capitalism and Society 3.
16. Stephan P (2012) How Economics Shapes Science: Harvard University Press.
17. Merton RK (1973) The sociology of science: Theoretical and empirical investigations. Chicago: University of Chicago Press. xxxi, 605 p. p.
18. Freeman R, Weinstein E, Marincola E, Rosenbaum J, Solomon F (2001) Competition and Careers in Biosciences. Science 294: 2293.
19. Haeussler C, Jiang L, Thursby J, Thursby M (2009) Specific and general information sharing among academic scientists. NBER Working Paper #15315.
20. Walsh JP, Cho C, Cohen W, M. (2005) View from the bench: Patents and material transfers. Science 309: 2002-2003.
21. Nelson R (2004) The market economy, and the scientific commons. Research Policy 33: 455-471.
22. Fleming L (2001) Recombinant uncertainty in technological search. Management Science 47: 117-132.
23. Uzzi B, Spiro J (2005) Collaboration and creativity: The small world problem. American Journal of Sociology 111: 447-504.
24. Hargadon A, Sutton RI (1997) Technology brokering and innovation in a product development firm. Administrative Science Quarterly: 716-749.
25. Lakhani K, Jeppesen L, Lohse P, Panetta J (2007) The Value of Openess in Scientific Problem Solving. Working Paper.
26. Lacetera N, Zirulia L (forthcoming) The economics of scientific misconduct. Journal of Law, Economics, and Organization.
27. Mulligan A (2005) Is peer review in crisis? Oral oncology 41: 135.
28. Raymond E (1999) The cathedral and the bazaar. Knowledge, Technology & Policy 12: 23-49.
29. Rennie D, Yank V, Emanuel L (1997) When authorship fails: A proposal to make contributors accountable. Journal of the American Medical Association 278: 579-580.
30. Sorenson O, Fleming L (2004) Science and the diffusion of knowledge. Research Policy 33: 1615-1634.
31. Dasgupta P, David PA (1994) Toward a new economics of science. Research Policy 23: 487-521.
32. Baldwin CY, Clark KB (2006) The architecture of participation: Does code architecture mitigate free riding in the open source development model? Management Science 52: 1116.
33. Franke N, Hippel E (2003) Satisfying heterogeneous user needs via innovation toolkits: the case of Apache security software. Research Policy 32: 1199-1215.

34. Lerner J, Tirole J (2002) Some simple economics of open source. Journal of Industrial Economics 50: 197-234.
35. O'Mahony S, Ferraro F (2007) The emergence of governance in an open source community. The Academy of Management Journal 50: 1079-1106.
36. Dahlander L, O'Mahony S (2010) Progressing to the center: Coordinating project work. Organization Science 22: 961-979.
37. Von Krogh G, Haefliger S, Spaeth S, Wallin MW (forthcoming) Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development. MIS Quarterly.
38. Hertel G, Niedner S, Herrmann S (2003) Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. Research Policy 32: 1159-1177.
39. Lakhani KR, von Hippel E (2003) How open source software works: "free" user-to-user assistance. Research Policy 32: 923-943.
40. Sauermann H, Cohen W (2010) What makes them tick? Employee motives and industrial innovation. Management Science 56: 2134-2153.
41. Hars A, Ou S (2002) Working for Free? Motivations for Participating in Open-Source Projects. International Journal of Electronic Commerce 6: 25-39.
42. Shapin S (2008) The scientific life: a moral history of a late modern vocation: University of Chicago Press.
43. Lakhani K, Wolf R (2006) Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In: Feller J, Fintzgerald B, Hissam S, Lakhani K, editors. Perspectives on Free and Open Source Software: MIT Press.
44. David PA, Shapiro JS (2008) Community-based production of open-source software: What do we know about the developers who participate? Information Economics and Policy 20: 364-398.
45. Roberts J, Hann IH, Slaughter S (2006) Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects. Management Science 52: 984-999.
46. Henkel J (2009) Champions of revealing—the role of open source developers in commercial firms. Industrial and Corporate Change 18: 435-471.
47. Von Hippel E, Von Krogh G (2003) Open source software and the" private-collective" innovation model: Issues for organization science. Organization Science: 209-223.
48. Bessen J (2006) Open source software: Free provision of complex public goods. In: Bitzer J, Schröder PJH, editors. The economics of open source software development: Elsevier. pp. 57.
49. Corbet J (2012) The Kernel Report. Linux Conference. Ballarat (Australia).
50. Bonaccorsi A, Rossi C, Giannangeli S (2006) Adaptive entry strategies under dominant standards: Hybrid business models in the Open Source software industry. Management Science 52: 1085-1098.
51. Wicks P, Vaughan TE, Massagli MP, Heywood J (2011) Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. Nature Biotechnology 29: 411-414.
52. Henkel J (2006) Selective revealing in open innovation processes: The case of embedded Linux. Research Policy 35: 953-969.
53. McGowan D (2001) Legal implications of open-source software. University of Illinois Law Review: 241.
54. Heller M, Eisenberg R (1998) Can patents deter innovation? The anticommons in biomedical research. Science 280: 698-701.
55. Kuhn T, S. (1996) The Structure of Scientific Revolutions: University of Chicago Press.
56. Owen-Smith J (2001) Managing laboratory work through skepticism: Processes of evaluation and control. American Sociological Review 66: 427-452.