SISSA – International School for Advanced Studies Journal of Science Communication
ISSN 1824 – 2049 http://jcom.sissa.it/

RECEIVED: September 17, 2009
PUBLISHED: March 22, 2010

Article

SPECIAL ISSUE ON PEER-TO-PEER AND USER-LED SCIENCE

Open science: policy implications for the evolving phenomenon of user-led scientific innovation

Victoria Stodden

ABSTRACT: From contributions of astronomy data and DNA sequences to disease treatment research, scientific activity by non-scientists is a real and emergent phenomenon, and raising policy questions. This involvement in science can be understood as an issue of access to publications, code, and data that facilitates public engagement in the research process, thus appropriate policy to support the associated welfare enhancing benefits is essential. Current legal barriers to citizen participation can be alleviated by scientists' use of the "Reproducible Research Standard," thus making the literature, data, and code associated with scientific results accessible. The enterprise of science is undergoing deep and fundamental changes, particularly in how scientists obtain results and share their work: the promise of open research dissemination held by the Internet is gradually being fulfilled by scientists. Contributions to science from beyond the ivory tower are forcing a rethinking of traditional models of knowledge generation, evaluation, and communication. The notion of a scientific "peer" is blurred with the advent of lay contributions to science raising questions regarding the concepts of peer-review and recognition. New collaborative models are emerging around both open scientific software and the generation of scientific discoveries that bear a similarity to open innovation models in other settings. Public engagement in science can be understood as an issue of access to knowledge for public involvement in the research process, facilitated by appropriate policy to support the welfare enhancing benefits deriving from citizen-science.

Introduction

"A few years ago when Tom Ray, a biologist turned nerd, created a digital habitat in a small computer and then loosed simple digital organisms in it to procreate, mutate, and evolve, he was no longer merely modeling evolution or collecting data. Instead, Ray had created a wholly new and novel example of real evolution. That's nerd science. As models and networked simulations take on further complexity and presence, their role in science will likewise expand and the influence of their nerd creators increase."[1] Author Kevin Kelly is describing a scientist who turns to using computational methods, thereby becoming a "nerd scientist." Is it really that much more difficult, or difficult at all, for non-scientists to do the same? Kelly goes on to state that "third culture meant a streetwise science culture, one where working scientists communicated directly with lay people, and the lay challenged them back. This was a peerage culture, a peerage that network technology encouraged." Massive increases in the pervasiveness of computation in scientific research are resulting in the widespread digitization of science. Not only are datasets and code now common in research, but the steps computational scientists make to obtain their results can be recorded digitally, capturing the sequence of decisions necessary to generate and re-generate discoveries, often down to the very last precise detail. At the same time, the advent of the Internet provides a mechanism for transmission of digitally encoded data, such as research findings, scientific articles, and supporting code and data, creating the potential for a world where not only are scientific results disseminated to the public, but so is the underlying reasoning, all ready for replication, testing, and experimentation by others.

This article surveys the current state of user-led science actively underway on the Web in Part 1, contrasts it with the ideal of full intellectual engagement described above and suggests necessary public policy measures in Part 2. Part 3 discusses the impact of the incipient involvement of non-scientists on scientific innovation.

1. The emergent phenomena of citizen-science

Since the 17th century scholarly communication has been carried out via articles in written periodical journals.[2] The Internet is a new mechanism for researchers to post the digital aspects of their work for public download, changing the centuries old way scientists have publicized and communicated their work.[3] This change has been made manifest in academia today primarily as an *additional* aspect to traditional publishing. Computational scientists can choose to make their papers, code, and data available publicly on the web, and many do.[4]

Access to scientific knowledge is also changing as ever-cheaper computing power and disk space have not only made more efficient tools available for research in the tradition of the scientific method, but have also digitized an increasingly large proportion of scientific research. An example chosen from statistics, the field that underlies data analysis and inference, illustrates the point. Ten years ago, in the June 1996 issue of the Journal of the American Statistical Association (JASA), 9 of the 20 articles published were computational, and none of those released their code or mentioned what software package they had used to produce their results.[5] In the June 2006 issue of JASA, 33 of 35 articles were computational, with only four mentioning the software package they had used, and three making their code publicly available (the fourth indicated the code was available upon request). This demonstrates that both the rise in digitization of scientific research and the proliferation of Internet use makes an ideal setting, although not yet fully utilized, for the public communication of science including the published results, as well as any code or data that underlie them. This is an important change and an apparent invitation to computational and data-driven citizen-science. The implications for how we as a society understand and learn about our world are profound.

With the data and the complete set of instructions for generating the results, in theory anyone with internet access and an appropriate computing platform can reproduce results, tweak them, rerun scripts, modify algorithms, try the algorithm on new data, and potentially contribute new scientific discoveries.[6]

Citizen-science on the web

It is becoming increasingly routine for computational scientists to make their underlying code and data available when publishing their results, but this is far from a typical behavior, and the scenario described above has yet to become commonplace.[7] There is reason to believe it will do so, and interested members of the general public will be able to avail themselves of greater opportunities to contribute to the process of scientific discovery. Innovation in the commercial setting has been shown to derive in part from the users of the products themselves, and there is reason to think this type of open innovation will extend to the scientific context as well.[8]

The most common way the general public interacts with scientific research today is by collecting data, and this can perhaps be considered the first step in wider scientific engagement. Some scientific researchers have realized they can generate new datasets for study by asking for help from people who may be better located to perform the data collection.[9] Birdsource hosts the Great Backyard Bird Count and counts species across America.[10] People are called to help scientists track plant and animal life cycle stages at the USA National Phenology Network (USA-NPN).[11] "The program is designed for people interested in participating in climate change science, not just reading about it," said Jake Weltzin, executive director of the USA-NPN and a scientist at the U.S. Geological Survey.[12] Inaturalist.org allows volunteers to add a pin in a Google map every time they see an animal or plant of note. The site records what was sited as well as its geographic coordinates and time of sighting. The Great Sunflower Project, lead by researchers at San Francisco State University, encourages people to "join the hunt for bees!"[13] "By watching and recording the bees at sunflowers in your garden, you can help us understand the challenges that bees are facing."[14] Their goal is to understand urban pollinators and the role they have in providing foodsources to people.

The Community Collaborative Rain, Hail, and Snow Network collects volunteer reports of precipitation across the United States, creating datasets and visualizing the data.[15] The Maryland Science Center allows individuals to upload UV data readings and compare them to instrumented readings and predictions.[16] This is done using the Volksdata tool that facilitates data upload to combine "the power of social networking with custom web applications to create your own distributed science project."[17]

Websites have appeared that call for disease and health monitoring. Patientslikeme.org is a collaborative treatment data-sharing platform that facilitates community formation and medication tracking for a variety of diseases such as ALS, HIV/AIDS, fibromyalgia, MS, Parkinson's disease, various mood conditions, and several very rare diseases. Patients report on how they feel and what drugs seem to be working in order to learn successful treatment strategies by viewing others' data and their own. These data are also useful for drug research in pharmaceutical companies. Lybba.org has the goal of making evidence-based medical information available and understandable on the web, building communities for discussion and shared experiences: to "improve the world's medical knowledge and improve people's relationship to medicine into the future."[18] User-led science can more closely address users' needs, particularly with respect to medical disease research, since users may have more information in some cases than scientists. The for-profit company 23andMe is gathering individuals' genetic information to both inform that individual as well as create datasets to enable genetic research on groups with particular characteristics of interest. The website DIYbio.org "aims to help make biology a worthwhile pursuit for citizen scientists, amateur biologists, and DIY biological engineers who value openness and safety. This will require mechanisms for amateurs to increase their knowledge and skills, access to a community of experts, the development of a code of ethics, responsible oversight, and leadership on issues that are unique to doing biology outside of traditional professional settings."[19] Participants are interested in bio- and genetic engineering, often calling themselves "biohackers," and maintain an open community.[20]

The Open Dinosaur Project engages nonscientists to glean dinosaur bone measurements from published articles and build comprehensive datasets studying evolution and other trends.[21] The results are intended for publication, including contributing citizen-scientists as co-authors. On their webpage the scientists state their motivation: "Anyone can do science – this firm belief is part of why we started the Open Dinosaur Project." They have built an active community of nearly 50 serious contributors.

Further inroads into participation by nonscientists in areas where typically only scientists would operate occurs in astronomy. Amateur sky watchers upload images, identify new features, and are publishing papers in academic astronomy journals (typically with an academic co-author). Galaxy Zoo is a website that allows image upload and labeling to occur, by anyone and new discoveries have been contributed by amateurs.[22] "A team of astronomers has discovered a group of rare galaxies called the 'Green Peas' with the help of citizen scientists working through an online project called Galaxy Zoo. The finding could lend unique insights into how galaxies form stars in the early universe."[23] Additionally, in September 2009 a high-school student named Lucas Bolyard discovered a rare astronomical object.[24] Finally, the Stardust@Home project is an "interactive Internet-based search for interstellar dust" in the sample collected by the Stardust spacecraft.[25] That search job would be prohibitively time consuming for a single lab.

Opportunities for lay engagement in science are collected at scienceforcitizens.net with over 100 projects and more being added each week.[26]

2. Public policy and lay participation

In the examples in the previous section, the volunteer is typically being directed by the expert scientist and has very little opportunity to contribute in a more deeply intellectual way (exceptions are the identification of new space artifacts, biohacking, and the relating of detailed information beyond data recording, as in the health examples). Public involvement in science at this stage appears to be largely as observer and recorder. It seems the predominant mode of citizen involvement is as an instrument of the scientist, rather than as a leader and originator of the research. As scientists share more of their research on the web, these scientist-driven research models may recede as future web-surfers find a greater selection of digital and computational problems to engage with. The democratization of science cannot simply mean the direction of citizens at the scientist's behest since intellectual engagement, particularly in the form of framing and solving problems, is at the heart of scientific discovery. The engagement of non-scientists in research as facilitated by the Internet is very new and it is reasonable to expect it to evolve toward deeper engagement of the non-professional especially as opportunities increase. Across the sciences, throughout government agencies, in engineering and in business there are major efforts underway to collect and analyze enormous amounts of data into vast databases. Gone are the days of competition over a small number of closely guarded meager datasets, and now the sheer number of scientific problem accessible to people with an internet connection and perhaps some computing power extends an invitation to those outside the ivory tower to participate in scientific discourse.

Scientific questions increasingly pervade the suite of public policy problems voters must address. Regulatory responses to some of the most salient issues today – climate change, global warming, medicine and health care policy for example – rest fundamentally on data and scientific analysis.[27] Results suggest that health care research that lacks citizen input tends to focus heavily on disease causes as opposed to patient decision making with regard to treatment. [28] The recent move by the Whitehouse towards "evidence-based policy" reflects the scientific nature at the core of today's issues.[29] In fact, much government funded research is accordingly mandated to be open to the public, but this is not always enforced primarily for legacy reasons.[30]

Traditional empirical research holds strict standards to assure peers that all efforts have been taken to root our error: the use of the statistical hypothesis testing framework, controlled experiments, and the generation of routinely verifiable knowledge through standardized reproducibility information given in publications (data, materials, and methods). With the advent of computational research a new, and as yet seldom recognized, image of the scientist is emerging: that of "a computer jockey working at all hours to launch experiments on computer servers.[31] The communication of computational results has not yet caught up to the standards for communication of traditional empirical results. To achieve the same standards computational scientists must generate *routinely verifiable knowledge*. To do this, release of the data and code underlying their results must become accepted and expected practice. Citizen-science benefits from the availability of data, code, results and state of the art knowledge of scientific discoveries. In the U.S. copyright law establishing exclusive rights for authors of original works of creative expression by default and this encompasses much scientific output, the manuscript, figures, tables, code, and the selection and arrangment of the data.[32] This creates a barrier to the sharing of scientific scholarship, limiting the ability of others to copy, use, build upon, or alter the research, especially in the case of computational work. The "Reproducible Research Standard" (RRS) is an intellectual property framework designed to encourage and facilitate code and data release, alongside the publication of results.[33] The RRS suggests releasing all aspect of the research with licensing terms that permit others to copy and build on the work while giving attribution to original authors, in the spirit of longstanding scientific norms.[34]

Using an open lincencing structure such as the Reproducible Research Standard on all components of scientific scholarship will not only promote integrity in computational science by facilitating reproducible scientific investigation and encouraging greater collaboration, but also give rise to the engagement of the larger community in scientific learning and discovery. Recent research indicates scientists are willing to freely reveal information about their work that citizen scientists can draw upon in their research.[35] At the moment a small proportion of computational research is fully open, and wider engagement in this area depends on the availability of the code and data.

User-led innovation in industry is a well-studied field. A number of studies have found that a substantial proportion of new product innovations in industry derive from users, rather than within-company research and development.[36] Insofar as the industrial setting resembles the scientific research setting, these findings can guide understanding of innovation in science, thinking of the general public as the "users" or scientific knowledge. A body of literature addresses the social welfare properties of user innovation in industry and argues that a world where both manufacturers and users innovate is likely to be better off than one in which only manufacturers innovate.[37] It is worth considering whether the involvement of volunteers without scientific training would create a net benefit for science. Whether a result is accepted as a contribution to society's stock of knowledge is decided by the discoverer's peers, traditionally other scientists engaged in research in the relevant area acting as reviewers of the work. A proliferation of amateur results would tax the existing system and call into question what is meant by a peer in scientific research. Open questions remain: would peer-produced science by judged by others (broadly-defined) involved in scientific research? Would citizen-scientists be held to the same standards as trained scientists? Typically scientists review articles submitted for publication to determine their scientific worthiness and shake out where they fit in the prestige hierarchy of journals. It is plausible to imagine this practice becoming untenable, through sheer volume alone, if citizen-scientist articles were added to the pool in any kind of number. Since lay scientists have tended to focus on areas where little reward has been traditionally given, such as data collection, this has not yet been an issue.[38] Scientific research on a particular problem also tends to be a very long discussion, involving numerous papers, many studies, and often taking place over the span of years. Excerpting a small slice of this discourse can produce misunderstandings of the larger research question. Open release of computational science provides a unique setting for the wider understanding of scientific research at a level not previously possible but questions remain about the evolving role of gatekeepers in the designation and recognition of scientific facts.

3. Reflections on scientific innovation models

At core in this discussion is the issue of scientific progress. Understanding of our world is thought to be a public good, and a set of norms governing scientific behavior has evolved to create incentives that facilitate a tradeoff between private gain and adding to society's stock of knowledge. In 1942 Merton studied mechanisms by which society's stock of scientific knowledge is created, defining a scientist as one who follows "the ethos of science." This ethos was comprised of four norms, in particular the eschewing of property rights, with the exception of naming discoveries, in exchange for recognition and esteem.[39]

Citizen-science bends this incentive model to one closer to that of open source software. Citizen-scientists do not seem to be seeking recognition and esteem from the scientific community through their participation, but appear to enjoy discovery and making meaningful contributions to research for its own sake.[40] Although scientists surely enjoy discovery, incentives are different in traditional science, where citation is the currency of success, suggesting that lay contributions may not seeks the same recognition as is traditionally the case. This opens the door to a potential collaboration model between scientists and engaged non-scientists, aside from the one described previously, with non-scientists acting as assistants to scientists particularly in the collection of data. Code is a becoming central to scientific progress as methods are increasingly encoded in digital form and tools of analysis are established for data. As in the open source community, as scientific software is increasingly released on the web, communities could emerge surrounding the development and maintenance of this software, similar to that surrounding the Linux kernel for example. This would be a valuable contribution to science, and one not typically filled by scientists today.

A more subtle concern is at work here: code and data are relatively new elements of the scientific effort and are becoming fundamental to scientific advancement – sharing these elements is crucial for the generation of routinely verifiable knowledge. With the new importance of these crucial elements of research, questions are emerging that prompt discussion of whether they should earn a higher recognition value. Data gathering is distinct from solutions and discoveries, but at the moment there is no standard citation mechanism for code and data as there is for published ideas and discoveries.[41] Elevating the value of data collection and collation and code clarity and effectiveness could be an important step in encouraging code and data release. Citizen-science appears to complement professional science as non-scientists are able to contribute data to research that might not have been available otherwise, as well as encouraging traditional scientists to release the code and data associated with their published results.

In the case of Tom Ray, the biologist turned "nerd scientist" at the opening of this article, he carried with him deeply embedded norms and a structured way of asking questions and seeking answers from his training as a biologist. A scientist is distinguished as someone who practices a particular set of norms in the pursuit of knowledge and without an understanding of these norms it is exceedingly difficult to join the scientific dialog. I believe that the opportunity for interested non-scientists to understand and manipulate experimental results – a scenario that can result from truly reproducible research – will create an exploratory ground for the understanding of these norms beyond academia and the ivory tower.

Opportunities for contributions to academic science are increasingly occurring outside traditional institutions. In physics the fully public website arXiv.org and myriad open blogs by scientists allow individuals not affiliated with an academic institution, such as well-known physicist Garrett Lisi, to engage and change the state of the art in scientific research.[42] ArXiv.org has only a very light screening for publicly posting submitted articles and is the primary mechanism for research communication in physics. Other fields have begun to use arXiv.org as well – mathematics, computational biology, and computer sciences articles are appearing on the site. All these articles are available to anyone who accesses the website.[43] Recently, a mathematical proof was collaboratively obtained using an open blogging platform, outside traditional academic research. Timothy Gowers, professor of mathematics at Cambridge University, launched the PolyMath Project, an effort in massively collaborative theorem proving. In its first use, a proof was found via group discussion for a longstanding theorem in a short period of time.[44] The contributors were all established mathematicians, perhaps necessary for the level of mathematics involved, but the collaboration was done openly on the Web with the intention that anyone could participate, thus serving as a model for other proofs and scientific discourse.

Conclusion

In 1994 Sharon Terry's two children were diagnosed with a rare disease, PXE. With no medical training she turned from her work as college chaplain to a career as an advocate for open access to scientific information and as a citizen-scientist by founding PXE International to support scientific research into PXE, and being involved in the cloning of the gene that is mutated in PXE patients.[45] Terry worked to isolate the DNA and sequence the responsible gene and is listed a co-inventor on the resulting patent. She has helped build a consortium of 19 research laboratories to develop diagnostic tests for PXE, directly using her scientific results.[46] This has been done without internet access to medical literature (she paid fees to visit medical libraries in person) and opens the question of what will be possible with Web access not only to scientific results, but to the underlying code and data thereby allowing people like Terry to contribute to computational science.

The enterprise of science is undergoing fundamental changes, particularly with regard to the increasing pervasiveness and scale of computing, having implications for how scientists obtain results and share their work. Computational scientists can be encouraged to release code and data that underlie their published results, affording anyone with internet access the opportunity to inspect the full set of instructions that generated the results. If the web-surfer wishes to run the code, he or she will be able to manipulate the code, change parameters or datasets, and explore the results. Policy responses such as the Reproducible Research Standard can help bring greater transparency to computational science, encouraging scientific involvement beyond data-collection and other supportive efforts.

Normative responses are equally important. Contributions from citizen-scientists put pressure on the very definition of scientific peer and thus on the practice of peer-review, aside from the potential increase in contributions. Further, current inconsistent citation practices for data and code reuse could be streamlined to create clarity in attribution and encourage greater sharing of code and data. Collaborative models between scientists and non-scientists can be formed and encouraged, perhaps in the spirit of the open source software movement as scientific software is increasingly released and reused. Public engagement in science can be understood as an issue of access to knowledge for public involvement in the research process, facilitated by appropriate policy to support the welfare enhancing benefits deriving from citizen-science.

Notes and references

[1] Quoted from K. Kelly (1998), *The Third Culture*, *Science* **279**(13): 992-993.

[2] J. Willinsky (2005), *The Access Principle: The Case for Open Access to Research and Scholarship*, MIT Press.

[3] C. Borgman (2007), *Scholarship in the Digital Age*, MIT Press.

[4] See V. Stodden, *The Scientific Method and Computation: Reproducibility in the Computational Sciences* (forthcoming), the goal of which is to understand why these scientists have made the decision to reveal or not reveal the code and data that underlies their published results.

[5] See V. Stodden (2006), *Model Selection When the Number of Variables Exceeds the Number of Observations*, Ph.D Thesis, Department of Statistics, Stanford University, available at http://www.stanford.edu/~vcs/thesis.pdf (last accessed Sept. 15, 2009).

[6] For a discussion of the moral foundations of access to knowledge, see J. Balkin, *What is Access to Knowledge*, http://balkin.blogspot.com/2006/04/what-is-access-to-knowledge.html (last accessed Feb. 25, 2010).

[7] For example, on September 14, 2009 five universities signed a compact committing to "the timely establishment of durable mechanisms for underwriting reasonable publication charges for articles written by its faculty and published in fee-based open-access journals and for which other institutions would not be expected to provide funds." From http://www.oacompact.org/compact/. See also http://hul.harvard.edu/news/2009_0914_compact.html, last accessed Sept. 15, 2009).

[8] See E. von Hippel (1976), *The Dominant Role of Users in the Scientific Instrument Innovation Process*, *Research Policy* **5**: 212-39; E. von Hippel (1977), *The Dominant Role of Users in the Semiconductor and Electronic Subassembly Process Innovation*, *IEEE Transactions on Engineering Management* **EM-24**(2May): 60-71; and E. von Hippel (1988), *The Sources of Innovation*, New York: Oxford University Press.

[9] This is generally termed 'crowdsourcing' and is in regular use beyond scientific applications. See J. Howe (2006), *The Rise of Crowdsourcing*, *Wired* Issue 14, available at http://www.wired.com/wired/archive/14.06/crowds.html (last accessed Sept. 15, 2009).

[10] See http://www.birdsource.org (last accessed Sept. 15, 2009).

[11] See http://www.usanpn.org/ (last accessed Sept. 15, 2009).

[12] Quoted in quoted in http://latimesblogs.latimes.com/greenspace/2009/03/help-scientists.html (last accessed Sept. 14, 2009).

[13] See http://www.greatsunflower.org/ last (last accessed Sept. 15, 2009).

[14] Quotes from http://www.greatsunflower.org/ (last accessed Sept. 15, 2009).

[15] See http://www.cocorahs.org/ (last accessed Sept. 15, 2009).

[16] See http://marylandsciencecenter.org/exhibits/aura.html (last accessed Sept. 15, 2009).

[17] See http://www.volksdata.com/ (last accessed Sept. 15, 2009).

[18] Splash page of Lybba website. See http://www.lybba.org (last accessed Sept. 15, 2009), See also http://www.form.tv/vault/jd/lybba.mov (last accessed Sept. 15, 2009).

[19] See http://diybio.org/about/ (last accessed Feb 25, 2010).

[20] Some of the information used by the biohackers is available via http://openwetware.org/, a platform organized by MIT students to facilitate the open sharing of research information between engineering and biology labs. It has since grown into a standard tool used by many biology labs around the world to record details of their experiments in real time.

[21] See http://opendino.wordpress.com/faqs/ (last accessed Feb. 25, 2010).

[22] See http://www.galaxyzoo.org/ (last accessed Sept. 15, 2009).

[23] Reported in *Galaxy Zoo Hunters Help Astronomers Discover Rare 'Green Peas' Galaxies*, ScienceDaily, July 28, 2009. Available at http://www.sciencedaily.com/releases/2009/07/090727135527.htm (last accessed Sept. 15, 2009).

[24] See http://www.space.com/scienceastronomy/080925-student-discovery.html (last accessed Feb. 25, 2010). There is a long history of contributions by high school students in astronomy, see e.g., http://en.wikipedia.org/wiki/Clyde_Tombaugh (last accessed Feb. 25, 2010).

[25] See http://stardustathome.ssl.berkeley.edu/ (last accessed Sept. 15, 2009).

[26] See http://scienceforcitizens.net/ (last accessed Feb. 25, 2010). The website includes projects such as ChemSpider (verifying chemistry on the web) and the REEF Volunteer Fish Survey Project, for example.

[27] See V. Entwistle et al. (1998), *Lay perspectives: advantages for health research*, *British Medical Journal* **316**: 463-466. Available at http://www.bmj.com/cgi/content/full/316/7129/463?ijkey=fb17a26fd479ea64547c8bcc1eb52effa771e627&keytype2=tf_ipsecsha (last accessed Sept. 15, 2009).

[28] This suggests a role for user-led research in health care treatment options, precisely what the two health examples from the previous section, PatientsLikeMe.org and Lybba,org, claim as their motivation. See I. Chalmers (1995), *What do I want from health research and health researchers when I am a patient?*, *British Medical Journal* **310**: 1315-1318. Available at http://www.bmj.com/cgi/content/full/310/6990/1315?ijkey=ca7097f4cad46486c2726c5246828104dea30797&keytype2=tf_ipsecsha (last accessed Sept 15. 2009). See also A. Faulkner (2002), *User-led Research and Evidence-based Medicine*, *The British Journal of Psychiatry* **180**: 1-3. Available at http://bjp.rcpsych.org/cgi/content/full/180/1/1 (last accessed Sept. 15, 2009).

[29] See http://www.whitehouse.gov/omb/blog/09/06/08/BuildingRigorousEvidencetoDrivePolicy/ (last accessed Feb. 4, 2010).

[30] 38. Sharing of Findings, Data, and Other Research Products

a. NSF expects significant findings from research and education activities it supports to be promptly submitted for publication, with authorship that accurately reflects the contributions of those involved. It expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work. It also encourages grantees to share software and inventions or otherwise act to make the innovations they embody widely useful and usable.

National Science Foundation (NSF) Grant General Conditions (GC-1), June 1, 2007. Available at http://www.nsf.gov/pubs/policydocs/gc1_607.pdf (last accessed Sept. 4, 2007).

[31] V. Stodden et al. (2009), *Reproducible Research in Computational Harmonic Analysis*, *IEEE Computing in Science and Engineering* **11**(1), Jan/Feb,. Available at http://www2.computer.org/portal/web/csdl/magazines/cise;jsessionid=2C09BB6B29A7412E17AA8C3DE828F30E - 3 (last accessed Sept. 15, 2009).

[32] See Feist Publ'ns Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991) at 363-364.

[33] For details see V. Stodden (2009), *Enabling Reproducible Research: Licensing for Scientific Innovation*, *International Journal of Communications Law and Policy*, Issue 13, Winter 2009. Available at http://www.ijclp.net/issue_13.html (last accessed Sept. 15, 2009).

[34] The entire compendium, including paper, code, and data, is available on the Internet, and each is subject to attribution-only licensing or public domain certification. For more detail see V. Stodden (2009), *Enabling Reproducible Research, Licensing for Scientific Innovation*, *International Journal of Communications Law and Policy*, Issue 13, Winter 2009. Available at http://www.ijclp.net/issue_13.html (last accessed Sept. 13, 2009).

[35] V. Stodden, *The Scientific Method in Practice: Reproducibility in the Computational Sciences*, forthcoming, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1550193 (last accessed Feb. 25, 2010).

[36] See E. von Hippel (1976), *The Dominant Role of Users in the Scientific Instrument Innovation Process*, *Research Policy* **5**: 212-239.

[37] See e.g. J. Henkel and E. von Hippel (2005), *Welfare Implications of User Innovation*, *Journal of Technology Transfer* **30**(1-2): 73-87.

[38] There is reason to believe increased credit for data generation and curation is ascendant as science becomes increasingly digitized, as calls for data citation and increased scientific credit are being made. See e.g. M. Altman and G. King, *A Proposed Standard for the Scholarly Citation of Quantitative Data*, Available at http://www.dlib.org/dlib/march07/altman/03altman.html (last accessed Feb. 25, 2010). The public availability of research articles can be thought of as data itself, see P. Bourne, J. Fink, and M. Gerstein (2008), *Open Access: Taking Advantage of the Full Content*, *PLoS Comput Biol* **4**(3), available at http://www.ploscompbiol.org/article/info:doi%2F10.1371%2Fjournal.pcbi.1000037 (last accessed Feb 25, 2010).

[39] R.K. Merton (1942), *The Normative Structure of Science*.

[40] See K. Lakhani and R. Wolf (2005), *Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects*, In *Perspectives on Free and Open Source Software*, edited by J. Feller, B. Fitzgerald, S. Hissam, and K. R. Lakhani (MIT Press). They found that many programmers contributed to open source projects simply because they enjoyed learning from the coding experience.

[41] For a suggestion see M. Altman and G. King (2007), *A Proposed Standard for the Scholarly Citation of Quantitative Data*, *D-Lib Magazine* **13**(3/4 March/April). Available at http://dlib.org/dlib/march07/altman/03altman.html (last accessed Sept. 15, 2009)

[42] See Lisi's open research website at http://deferentialgeometry.org/ and his talk at http://www.ted.com/talks/garrett_lisi_on_his_theory_of_everything.html (last accessed Sept. 15, 2009).

[43] Similar websites are available for other fields, such as http://www.ssrn.com, and journals that fall under the label of Open Access, such as the Public Library of Science, PubMed Central, and many others. Many computational scientists self-archive their papers on their academic websites and are available for download.

[44] See http://gowers.wordpress.com/2009/01/27/is-massively-collaborative-mathematics-possible/ (last accessed Sept 15, 2009).

[45] See *Patient Advocate Calls for Open Access*, *Open Access Now*, October 4, 2004. Available at http://www.biomedcentral.com/openaccess/archive/?page=features&issue=21 (last accessed Sept. 15, 2009).

[46] See also T. Foremski (2009), *How a Software Engineer Tried to Save his Sister and Invented a Breakthrough Medical Device*, *ZDNet* Sept. 2, available at http://blogs.zdnet.com/Foremski/?p=760 (last accessed Sept. 15, 2009). Robert Goldman researched and created an FDA–approved medical device without any medical training.

Author

Victoria Stodden is Postdoctoral Associate in Law and Kauffman Fellow in Law in the Information Society Project at Yale Law School, and a Fellow at Science Commons. She has a Ph.D. in statistics from Stanford University and earned an M.L.S. from Stanford Law School. She has completed postdoctoral appointments at the Sloan School of Management at MIT and the Berkman Center at Harvard Law School. She can be reached at vcs@stanford.edu and she blogs at http://blog.stodden.net. E-mail: vcs@stanford.edu.

HOW TO CITE: V. Stodden, *Open science: policy implications for the evolving phenomenon of user-led scientific innovation*, *Jcom* **09**(01) (2010) A05.