Engaging Science, Technology, and Society 3 (2017), 560-574 DOI:10.17351/ests2017.124

Crowdsourcing Undone Science

GWEN OTTINGER[1]
DREXEL UNIVERSITY

Abstract

Could crowdsourcing be a way to get undone science done? Could grassroots groups enlist volunteers to help make sense of large amounts of otherwise unanalyzed data—an approach that has been gaining popularity among natural scientists? This paper assesses the viability of this technique for creating new knowledge about the local effects of petrochemicals, by examining three recent experiments in crowdsourcing led by non-profits and grassroots groups. These case studies suggest that undertaking a crowdsourcing project requires significant resources, including technological infrastructures that smaller or more informal groups may find it difficult to provide. They also indicate that crowdsourcing will be most successful when the questions of grassroots groups line up fairly well with existing scientific frameworks. The paper concludes that further experimentation in crowdsourcing is warranted, at least in cases where adequate resources and interpretive frameworks are available, and that further investment in technological infrastructures for data analysis is needed.

Keywords

citizen science; crowdsourcing; big data; community-based participatory research; undone science; environmental justice

Crowdsourcing Undone Science

Could crowdsourcing—specifically, distributing data analysis tasks to volunteers—be a way to get undone science done? Natural scientists have been enrolling volunteers in projects like Galaxy Zoo and eMammal as a way of extending their data processing power (Bonney et al. 2016;

---

[1] Gwen Ottinger, Email: ottinger@drexel.edu

McShea et al. 2016). Could grassroots groups, confronting the fact that scientific research tends to focus on questions of interest to elites at the expense of those relevant to marginalized communities (Hess 2007; Frickel et al. 2010), adopt the same strategy? Might they recruit relatively well-off science enthusiasts (Pandya 2012) to help create knowledge in areas where science has been left undone by the scientific establishment but that are of great concern to the (typically much less well-off) communities directly affected by fossil fuel extraction, energy generation, and petrochemical production?

Thus far, few groups concerned about industry's environmental impacts have attempted to crowdsource data analysis—even as more and more of them involve volunteers and/or local activists in generating data with low-cost, user-friendly sensors (e.g. Jalbert and Kinchy 2016). But a few grassroots experiments in distributed data analysis, examined in light of research on scientist-led crowdsourcing projects, suggest two provisional answers to the question of whether crowdsourcing can be a useful method for transforming what science gets done, and on whose behalf. First, the organizational and technological resources required to run a volunteer crowdsourcing program may be prohibitive for some grassroots groups—although existing networks of galvanized community members may help to make up for a lack of professional staff, for example. Second, crowdsourcing is likely to work best in situations where science is "undone" in the sense that relatively clear analytical frameworks are not being applied to existing data. On the other hand, where grassroots groups are unsatisfied with standard scientific frameworks for interpreting data and struggling to articulate alternatives (see Ottinger 2017), they are unlikely to be able to create the well-structured, low-complexity data analysis tasks needed to enroll large numbers of volunteers.

Aspects of Undone Science

The idea of "undone science" captures the fact that scientific research is not evenly distributed across possible areas of inquiry; rather, areas that interest political and economic elites are far more likely to be investigated than are questions that concern marginalized populations (Hess 2007). Among the areas that have been systematically under-investigated by mainstream science are the environmental and health impacts of industrial pollution (Frickel et al. 2010). By engaging in knowledge production of their own, "fenceline communities" and environmental health and justice non-profits attempt to do science that university, regulatory, and industry scientists have neglected, partially in the hopes that greater knowledge will amplify their political demands, including for government intervention to protect public health (Kinchy et al. 2014).

The problem of undone science is not, however, reducible to a simple absence of inquiry. Dynamics other than straightforward inattention may result in certain kinds of knowledge not being pursued (Proctor 2008). With respect to crowdsourcing, two in particular are worth noting: inadequate scientific frameworks, and overwhelming amounts of information. Some broad areas —e.g. air quality—receive a great deal of attention from researchers, but scientists frame their research in a way that fails to address questions of concern or to map on to local knowledge. Scientists investigate the chronic and acute effects of toxic air pollutants, for instance, without inquiring into the long-term health effects of repeated bursts of exposure to sub-acute levels (Ottinger 2010). Undone science in these cases is not a product of outright inattention to an issue —as it often is—but of framing the issue in a way that leaves consequential phenomena unexplored.

Similarly, undone science often involves a paucity of data on which to base scientific claims—but not in every case. For example, in most places in the United States, including the neighborhoods closest to industrial facilities, ambient concentrations of toxic air pollutants are simply not monitored. The lack of monitoring arguably both constitutes undone science in itself and results in undone science around questions that would depend on that data. But in recent years, advances in sensing and internet technologies have drastically expanded the data available to researchers and the general public (Balaji Prabhu and Arpitha 2014), shifting the nexus of undone science. Now important research areas may remain un- or underexplored because of a lack of resources to turn plentiful data into meaningful knowledge (Ottinger and Zurer 2011), rather than from the absence of data. While information scarcity remains a problem in the era of "big data"—availability of and access to data are far from equitable (boyd and Crawford 2012)— information overload is also becoming a driver of undone science.

Experiments in Crowdsourcing

Crowdsourcing, in the sense of distributing data analysis tasks across a large number of volunteers,[2] is one strategy for dealing with information overload. In the natural sciences, it is most commonly used for image processing and other tasks on which humans outperform algorithms (Franzoni and Sauermann 2014), such as classifying galaxies by shape (Bonney et al.

[2] The need for specificity here derives from the fact that usage of the terminology is not consistent. Elsewhere in the literature, "crowdsourcing" may refer to distributed data collection, distributed data processing, or both, whereas distributed data processing may also be called "crowd science" or "online citizen science."

2016) or identifying animals caught in "camera traps" (McShea 2016). It has similar potential applications in the study of the effects of petrochemical pollution on communities, where grassroots groups' access to images and sensing data grows as technology advances and becomes less expensive.

Three recent projects used distributed data analysis approaches in their attempts to understand the impacts of the petrochemical industry: FrackFinder, the Shenango Channel, and Meaning from Monitoring, each of which I describe briefly below. FrackFinder, a project of the non-profit organization SkyTruth, enlisted volunteers to determine the nature and extent of natural gas extraction operations by examining aerial images—enabling them to show the expansion over several years of hydraulic fracturing in Pennsylvania and Ohio. I learned about FrackFinder through conversations and written exchanges with David Manthos during his tenure as SkyTruth's Communications Director, as part of preparing this article.

The Shenango Channel project was a collaboration between engineers from Carnegie Mellon University's Community Robotics, Education, and Technology Empowerment (CREATE) Lab and Allegheny County Clean Air Now (ACCAN). ACCAN was founded by citizens affected by DTE Shenango Coke, a facility outside of Pittsburgh that converted coal into coke, an essential ingredient in steel. Members of the group looked at footage of the plant and picked out instances of fugitive emissions or unplanned releases (Hsu et al. 2017). Together, they generated a collection of gifs of the smoke rising from the facility that an EPA official called "completely unacceptable" (Hopey 2015) shortly before DTE announced in December 2016 that it would shut down the facility (Moore 2015). I first learned of the Shenango Channel from Randy Sargent, researcher at the CREATE Lab and a collaborator on the Meaning from Monitoring project (described below), and have subsequently had extended conversations about the project with Randy and seven other participants, four from each of the organizations involved, about what made their collaboration a success.

Meaning from Monitoring is a participatory design project that I initiated[3] in the hopes of making real-time ambient air monitoring data more accessible and useful to the oil refinery-adjacent, San Francisco Bay area communities where the data are being collected.[4] After working with residents of Richmond, Crockett, Rodeo, and Benicia to design a new web interface—

[3] The project is funded by an award from the National Science Foundation (#1352143). Randy Sargent and Dawn Nafus, Intel Labs, are key collaborators on the project, and Drexel University undergraduate Amy Gottsegen programmed the website and led the data analysis effort described here.

[4] For more about the project, see "A Missing Link in Making Meaning from Air Monitoring" (*Backchannels*, April 16, 2016) or "Lessons Learned from an Experiment in Infrastructuring" (*Toxic News*, May 16, 2017).

enabling exploration of historical data for the first time—we found that we still did not know what the data meant. Drawing on Randy's experience in the Shenango Channel project, we asked members of our working group to identify "incidents," or problematic periods of high pollution, by examining and taking screenshots of graphed data. However, we were not able to transform the data into new knowledge claims about, or even a better understanding of, community air quality using this method, for reasons that will be discussed below.

Responding to Undone Science

The three projects vary in the manner in which the science they addressed was "undone." ACCAN and the CREATE Lab initially confronted an absence of data, until they set up a camera to take and upload a picture of the plant every five seconds. The images were stitched together automatically using CREATE Lab's TimeMachine software, but the two groups were still left with hours and hours of "footage" to sift through. FrackFinder and Meaning from Monitoring both started out with rich data sets that had not been analyzed. Although state government data on the extent of drilling operations was incomplete, FrackFinder was able to use high-resolution, publicly available aerial images from the US Department of Agriculture's National Agricultural Imagery Program. Meaning from Monitoring was initiated in response to the installation of six monitoring stations surrounding the Chevron refinery in Richmond, California, each of which detects concentrations of multiple chemicals every five minutes and reports them to a publicly available website, without any provisions for data analysis. The monitoring program was modeled after fenceline monitoring at the Philips 66 refinery in Rodeo, which has been on-going since 1996 without data from it contributing to knowledge of the refinery's effects on the community (Ottinger and Zurer 2011).

While the crowdsourcing aspects of all three projects ultimately responded to information overload rather than information scarcity, the three projects remained differently positioned with respect to scientific frameworks for interpreting the data. In both the FrackFinder and Shenango Channel cases, what project participants were trying to quantify mapped on to regulatory frameworks for understanding petrochemical impacts. Specifically, FrackFinder looked to document the size and extent of drilling operations; this is something that some states, including Pennsylvania, also claim to do. The FrackFinder project responded to the incompleteness of these records, but generated information commensurate with the state's existing records and recognizable to regulators. Similarly, participants in the Shenango Channel were examining images for smoke and other visible emissions, which environmental regulators recognize as an issue of concern: a number of participants had trained as "smoke readers," becoming certified in EPA Method 9 for "Visual determination of the opacity of emissions from stationary sources." While their project looked for and at smoke in a place that was not being systematically scrutinized by regulators, it did not have to contest regulatory ways of looking to do so.

In contrast, the Meaning from Monitoring project confronted science that was undone (from participants' perspective) because regulators were asking the wrong questions, not because they weren't asking questions at all. The data collected from real-time monitors were compared, on an on-going basis, to regulatory standards and community warning system levels that, if exceeded, trigger an alert to residents to shelter in place or evacuate. Criticisms of these levels are by now familiar: they are uncertain; they are not low enough; they don't account for potential synergistic effects of chemicals; they average concentrations over time periods that are too long (Tesh 2000; Ottinger 2010). With better access to real-time data, participants in Meaning from Monitoring hoped to make visible other patterns that they experienced, including sub-acute incidents, series of small releases leading up to larger ones, and worsening air quality when winds blew from the direction of the refinery. Because none of these phenomena was identified by regulators as a quantifiable issue of concern, however, project participants started out without a clear framework for identifying them in, or representing them with, the data they had.

Experiences from the three projects help to show how crowdsourcing techniques borrowed from the (mostly academic) natural sciences are likely to play out in situations where non-profit and community groups are trying to leverage large data sets as a way of addressing undone science. A central challenge for smaller groups will be to find the resources to create effective distributed data analysis projects. While this is an issue for all crowdsourcing projects, non-profit and grassroots groups may experience it more intensely—but may also be able to take advantage of human capital not normally available to scientists. A second challenge is presented by the mismatch of scientific frameworks and grassroots groups' questions that characterizes some cases of undone science. All crowdsourcing projects depend on the ability of project leaders to create discrete data analysis tasks for volunteers to complete; in cases where there are no ready-made metrics or markers of the phenomena grassroots groups wish to represent, it may not be possible to structure tasks that can be delegated to volunteers.

Resources for Crowdsourcing

A primary motivation—if not *the* primary motivation—for distributing data analysis tasks to volunteers is cost savings. Put simply, using unpaid labor can make it feasible to take on data sets that would be prohibitively costly to analyze if scientists, research assistants, or non-profit staff members had to be paid to do the data analysis. In a study of seven crowdsourcing projects, Sauermann and Franzoni (2015) estimated the value of volunteer labor as ranging from $22,000 for the smallest project to $650,000 for the largest.[5]

To garner these cost-savings, however, organizations must invest resources. Research on crowdsourcing projects makes clear that they require significant program management work, including structuring tasks and attracting volunteers (Franzoni and Sauermann 2014; Sauermann and Franzoni 2015). The latter especially can be problematic (and, one imagines, time consuming): most volunteers in crowdsourcing projects participate only once (Sauermann and Franzoni 2015), leaving project managers with the challenge of either boosting retention or constantly competing with other projects for new participants. While these investments in crowdsourcing projects have not been quantified, some research on volunteer involvement in data collection has also acknowledged the significant costs of running citizen science projects, especially in the start-up phases (Bonney et al. 2009; Danielsen et al. 2005). Effective crowdsourcing projects also depend on cyberinfrastructure, including software to enable volunteer contributions (see e.g. McShea et al. 2016), and Crall et al. (2010) point out that many small organizations running citizen science projects lack the resources to effectively manage and share their data electronically.

The three examples of crowdsourcing described above both underscore and help characterize the costs of distributing data analysis as a strategy for getting undone science done. Of the three projects, FrackFinder was the only one that used volunteers in the classic sense, and significant effort went into volunteer recruitment and management. They recruited participants online—an effort that was aided by a fortuitously timed *Washington Post* magazine article about one of SkyTruth's other projects—and garnered 223 volunteers for the project's first phase. As in other crowdsourcing projects, volunteer interest waned over time, so in the second phase of the project they partnered with college classes from several universities (a strategy shown to be successful in other citizen science projects; c.f. Dickerson-Lange et al. 2016) and organized events at which students came together over pizza to do the online classification work. Staff time was also spent structuring tasks that volunteers could do easily and confidently and setting up

[5] Based on average hourly wages for undergraduate research assistants. Values were similar but the range was smaller when the authors used Amazon's Mechanical Turk pay scale.

systems to ensure that volunteers didn't mistakenly identify shadows and other anomalies in images as fracking-related activity. Whether this investment saved SkyTruth time or money overall is an open question: according to former Communications Director David Manthos, the crowdsourcing approach used in the FrackFinder project did not necessarily result in its being completed faster or more effectively than had staff done the work themselves. Yet he still deemed it worthwhile because it represented an opportunity for citizen engagement and an investment in building capacity for future projects.

The Shenango Channel and Meaning from Monitoring projects, in contrast, distributed data analysis tasks to members of affected communities who were already involved in efforts to bring attention to and contest the impacts of petrochemical operations. Distribution of tasks was handled collectively in both cases: shared documents were set up to let participants claim time ranges to analyze, and results were posted to shared folders in the cloud. Recruitment and retention were less of an issue, since participants had more inherent motivation to complete data analysis tasks. Yet the success of the two projects depended in no small part on the level of organizing in involved communities. ACCAN, the group involved in the Shenango Channel project, was very active on multiple fronts, with half a dozen or more core members able to devote significant time and energy to organizing and collective efforts. These individuals both helped with the image analysis and recruited others to do so. In contrast, Meaning from Monitoring's core group was made up of a few participants from several communities affected by refinery pollution. While their individual communities were organized to varying extents, we were not able to translate that into broader participation in data analysis, because there was no other inter-community organization that could support the effort. As a result, participation was limited to a few members of the working group.

Participation in these two projects was also affected by the technological infrastructures involved, highlighting additional kinds of resources necessary for crowdsourcing undone science. A key breakthrough in the Shenango Channel project came when CREATE Lab engineers developed a way for participants to grab a chunk of footage showing a smoky plume or other release and save it as an animated gif. ACCAN members then circulated the gifs to each other, in addition to contributing them to the project's repository. The ability to make their work visible in this way—as opposed to, say, only being able to record the time stamps of observed emissions—could only have increased their motivation and sense of purpose.

In contrast, the method we chose for Meaning from Monitoring asked community participants to take and upload screenshots of graphs that showed incidents or abnormal levels of pollution. But taking a screenshot and navigating to graphs of specified date ranges proved challenging for some participants, despite support from one of the project's research assistants. Just as the technical affordances of the Shenango Channel project likely increased participation, the technical complexity of the Meaning from Monitoring project limited participation to the more technically confident of the core group.

Observations from these three experiments in distributed data analysis for undone science are largely in keeping with findings from research on scientist-led crowdsourcing and citizen science projects: project management requires resources; recruitment and retention is an on-going challenge; and cyberinfrastructures matter. Yet, from these experiments, we get a better idea of the particular shape that those issues may take in cases where crowdsourcing is attempted as a way of turning volunteer effort to the problems of communities affected by petrochemical pollution. In cases where community organizing is strong, project management can be taken on by "volunteers" to some extent, and recruitment for data analysis projects can become just one more aspect of the outreach and mobilization that grassroots groups are already doing. The tradeoff, of course, is that such a strategy may limit the extent to which the labor of charitable outsiders can be turned to the cause of affected communities, and maintaining the engagement of the true volunteer has shown to be a challenge across all kinds of projects.

Further, these experiments suggest that technological infrastructures are likely to be an even greater challenge for grassroots crowdsourcing projects than for scientist-led ones. By their nature, distributed data analysis projects require a level of technological capacity on the part of the sponsoring organization that few community groups or small non-profits have. The CREATE Lab was easily able to provide that capacity for ACCAN in the Shenango Channel case. In Meaning from Monitoring, an interdisciplinary project led by a social scientist, providing a robust enough technological infrastructure has been an on-going challenge, and participation has suffered as a result.

Defining Tasks

The success of crowdsourcing depends, fundamentally, on being able to divide the work of data analysis into a series of discrete tasks that can be completed easily and accurately by volunteers with little or no training. Franzoni and Sauermann (2014) point to task complexity and task structure as defining characteristics of crowdsourcing projects, showing that the majority feature well structured tasks that are largely independent of one another (low complexity). Projects characterized by highly complex, ill-structured tasks do exist, they find, but they require participants to collaborate to a greater degree, building sequentially on each others' work and building a shared understanding of the overall problem over time. Such projects tend to limit participation to a smaller number of more skilled, more motivated participants.

Arguably, the likelihood of being able to break a data analysis project into low complexity, well structured tasks depends a great deal on the kinds of scientific questions being posed. Where the shape of the answer—the patterns or indicators one is looking for—is well understood, then discrete, granular tasks are easier to structure. For example, distinguishing house cats from foxes from raccoons in photos is a straightforward task, to the extent that mammalian species are relatively stable categories.

The issue of task structure and complexity becomes important for undone science to the extent that undone science varies in its relation to established scientific frameworks. Where science is undone in the sense that reasonably well defined frameworks are simply not being systematically applied to available data, creating low-complexity, well structured tasks is likely to be feasible. Leaders of the FrackFinder project, for example, knew what volunteers should be looking for, and were able to translate that into a simple task: in phase 1, volunteers were asked to classify images as showing evidence of active natural gas drilling, a well pad, or no well pad. In the second phase, they broke down the task of looking for fracking waste impoundments into two independent steps that could be completed by different volunteers, first identifying ponds in images, second classifying them as related to fracking or not.

In the Shenango Channel project, participants similarly knew what they were looking for: smoky emissions or other releases emanating from the Shenango coke works. Unlike FrackFinder, ACCAN members who contributed worked from a shared understanding of the problem rather than instructions from a project leader. Nonetheless, the task could have been conveyed to outsiders with a reasonably high rate of success—in the same way that the EPA is able to train people to be "smoke readers." Indeed, as ACCAN members started to home in on the kinds of images that represented the coke works' emissions problem, a graduate student from the CREATE Lab, Yen-Chia Hsu, was working on computer vision algorithms to automate the process of finding releases, though human involvement remained necessary to distinguish white smoke from clouds, for example (Hsu et al. 2016).

In contrast, where science is undone in the sense that the scientific questions being asked or frameworks being used are not adequate to representing a community's concerns or experiences, creating well-structured, low-complexity tasks may be a stumbling block for crowdsourcing. In the Meaning from Monitoring project, we knew we wanted to look for "incidents" in the data. Even starting from a notion that an incident was a time when pollution was abnormally high, we struggled to define an incident. For how long did chemical concentrations have to be elevated? How high? Did multiple chemicals have to be elevated simultaneously, or was one enough? Should we take into account only those that occurred when monitors were downwind of the refinery? We ended up with the kind of ill-structured problem that Franzoni and Sauermann (2014) describe as requiring a more sophisticated set of users to come to a collective understanding of the problem space. Being residents of the affected area, with both local knowledge and intrinsic motivation, our working group would perhaps be well positioned to take on this kind of task, if we were to set it up in a way to enable more active collaboration, rather than assuming that the work could be done piecemeal. However, the ill-structured nature of the project would likely make it unsuitable for a broad base of uninvolved volunteers, of the sort that FrackFinder enlisted. A crowdsourcing approach thus seems relatively difficult to apply to undone science in cases where standard scientific frameworks or questions are being challenged, and new ways of understanding data must be invented.

Pursuing Environmental Justice through Crowdsourcing?

Can crowdsourcing help get undone science done—and disrupt, if modestly, the normal power relations around science—by distributing data analysis tasks to science hobbyists in the service of communities affected by petrochemicals and other hazardous industry? The three experiments discussed here suggest that crowdsourcing could have potential as a new mode of knowledge production when certain conditions hold. Sponsoring groups must have the resources to recruit, retain, and organize the efforts of volunteers, as well as the resources to provide and maintain the technological infrastructures through which data analysis tasks are distributed and performed. In many circumstances, both informal groups of mobilized community members and more formal non-profit organizations could most likely provide the organizing resources. Providing the necessary technological infrastructures, however, is likely to require the resources of a technologically savvy non-profit or university partner.

These cases further suggest that crowdsourcing will be most effective as an approach to undone science where the scientific frameworks used to interpret data are not in question—or, at least, where project sponsors are clear about what the phenomenon that they are interested in capturing "look like" in the data. Where science is undone because scientists' ways of asking questions or looking at data are unsatisfying to affected communities, but alternative ways of looking remain elusive, distributed data analysis is unlikely to succeed. What is called for in these cases is a collaborative, open-ended process of looking at the data to better understand its relationship to local experience (Ottinger 2017). This could take the form of high complexity, low structure online citizen science of the sort that Franzoni and Sauermann (2014) describe, or instead be attempted through face-to-face workshops. Regardless, the local knowledge of affected communities will be a necessary element of any work to advance new scientific frameworks, making it difficult to delegate to uninvolved volunteers.

These initial findings, then, suggest the value of additional experiments that use crowdsourcing—specifically, distributed data analysis—as an approach to getting undone science done. Future experiments are likely to be most worthwhile in cases where data are abundant and interpretive frameworks that adequately represent grassroots groups' interests in the data are well developed.

At the same time, examining even a small number of existing experiments in crowdsourcing underscores the need for new kinds of investment in infrastructure. It is already well established that organizational capacity is necessary for grassroots groups to be effective participants in making new knowledge about petrochemical pollution and other local environmental impacts (e.g. Harrison 2011). And much effort and money has lately gone into supporting would-be citizen scientists in developing sensing capacity: witness, for example, the US EPA's "Air Sensor Toolbox for Citizen Scientists." However, the technical capacity and information infrastructures necessary for making sense of data have not received commensurate attention. For crowdsourcing to have a chance at addressing areas of undone science created by information overload, public agencies, foundations, and universities will need to invest in technical capacity building for distributed data analysis.

Author Biography

Gwen Ottinger is Associate Professor in the Department of Politics and the Center for Science, Technology, and Society at Drexel University. She is author of *Refining Expertise: How Responsible Engineers Subvert Environmental Justice Challenges*, which received the Rachel Carson Prize from the Society for Social Studies of Science in 2015.

Acknowledgements

The author wishes to thank David Manthos for numerous conversations about Frackfinder; Randy Sargent, Yen-Chia Hsu, JP Pachuta, Bea Dias, Leah Andrascik, Barb Pace, Thaddeus Popovich, and Angelo Taranto for sharing their experiences in the Shenango Channel project; and participants in the Meaning from Monitoring project's experiment in data analysis, including

Constance Beutel, Janet Callaghan, Kathy Kerridge, and Nancy Rieser. Research assistants Alex Boukedes, Amy Gottsegen, and Kristen Kepics each made a valuable contribution to some aspects of this article. Thank you also to Nick Shapiro, Max Liboiron, and Sara Wylie for their invitation to contribute to the thematic collection and their editorial input along the way, to an anonymous reviewer and ESTS editors for their comments on an earlier draft, and to the National Science Foundation for their support of the Meaning from Monitoring project (Award #1352143).

References

Balaji Prabhu, B. V., and S. Arpitha. 2014. "Study on Causes and Issues of Big Data." *International Journal of Engineering, Sciences and Research Technology* 3 (2):626-629.

Bonney, R., H. L. Ballard, R. Jordan, E. McCallie, T. B. Phillips, J. Shirk, and C. C. Wilderman. 2009. Public Participation in Scientific Research: Defining the field and assessing its potential for informal science education. In *A CAISE Inquiry Series Report*. Washington, DC: Center for Advancement of Informal Science Education (CAISE).

Bonney, R., C. B. Cooper, J. Dickinson, S. Kelling, T. B. Phillips, K. V. Rosenberg, and J. Shirk. 2009. "Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy." *BioScience* 59 (11):977-984.

Bonney, R., T. B. Phillips, H. L. Ballard, and J. W. Enck. 2016. "Can Citizen Science Enhance Public Understanding of Science?" *Public Understanding of Science* 25 (1):2-16.

boyd, d., and K. Crawford. 2012. "Critical Questions for Big Data." *Information, Communication, and Society* 15 (5):662-679.

Crall, A. W., G. J. Newman, C. S. Jarnevich, T. J. Stohlgren, D. M. Waller, and J. Graham. 2010. "Improving and Integrating Data on Invasive Species Collected by Citizen Scientists." *Biological Invasions* 12 (10):3419-3428.

Danielsen, F., N. D. Burgess, and A. Balmford. 2005. "Monitoring matters: Examining the Potential of locally-based approaches." *Biodiversity and Conservation* 14:2507-2542.

Dickerson-Lange, S. E., K. B. Eitel, L. Dorsey, T. E. Link, and J. D. Lundquist. 2016. "Challenges and Successes in Engaging Citizen Scientists to Observe Snow Cover: From Public Engagement to an Educational Collaboration." *Journal of Science Communication* 15 (1).

Franzoni, C., and H. Sauermann. 2014. "Crowd Science: The Organization of Scientific Research in Open Collaborative Projects." *Research Policy* 43 (1):1-20.

Frickel, S., S. Gibbon, J. Howard, J. Kempner, G. Ottinger, and D. J. Hess. 2010. "Undone Science: Charting Social Movement and Civil Society Challenges to Research Agenda Setting." *Science, Technology, and Human Values* 35 (4):444 - 473.

Harrison, J. 2011. "Parsing "Participation" in Action Research: Navigating the Challenges of Lay Involvement in Technically Complex Participatory Science Projects." *Society and Natural Resources* 24 (7):702-716.

Hess, D. J. 2007. *Alternative Pathways in Science and Industry: Activism, Innovation, and the Environment in an Era of Globalization*. Cambridge, MA: The MIT Press.

Hopey, D. 2015. "Regulators Reviewing Shenango Coke Works' compliance with 2012 Consent Decree." *Pittsburgh Post-Gazette*, November 19.

Hsu, Y., P. Dille, J. Cross, B. Dias, R. Sargent, and I. Nourbakhsh. 2017. "Community-Empowered Air Monitoring System." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, CO, pp. 1607-1619.

Hsu, Y., P. Dille, R. Sargent, and I. Nourbakhsh. 2016. *Industrial Smoke Detection and Visualization*. Pittsburgh, PA: The Robotics Institute, Carnegie Mellon University.

Jalbert, K., and A. Kinchy. 2016. "Sense and Influence: Environmental Monitoring Tools and the Power of Citizen Science " *Journal of Environmental Policy and Planning*.

Kinchy, A., K. Jalbert, and J. Lyons. 2014. "What is Volunteer Water Monitoring Good For?: Fracking and the Plural Logics of Participatory Science." *Political Power and Social Theory* 27:259-289.

McShea, W. J., T. Forrester, R. Costello, Z. He, and R. Kays. 2016. "Volunteer-run Cameras as Distributed Sensors for Macrosystem Mammal Research." *Landscape Ecology* 31:55-66.

Moore, D. 2015. "DTE to Close Shenango Plant, Affecting 173 Workers." *Pittsburgh Post-Gazette*, December 17.

Ottinger, G. 2010. "Buckets of Resistance: Standards and the Effectiveness of Citizen Science." *Science, Technology, and Human Values* 35 (2):244 - 270.

Ottinger, G. 2017. "Making Sense of Citizen Science: Stories as a Hermeneutic Resource." *Energy Research and Social Science*. DOI 10.1016/j.erss.2017.06.014

Ottinger, G, and R. Zurer. 2011. "Drowning in Data." *Issues in Science and Technology* 27 (3):71-73, 76-77, 80-82.

Pandya, R. E. 2012. "A Framework for Engaging Diverse Communities in Citizen Science in the U.S." *Frontiers in Ecology and the Environment* 10 (6):314-317.

Proctor, R. N. 2008. "Agnotology: A Missing Term to Describe the Cultural Production of Ignorance (and its Study)." In *Agnotology: The Making and Unmaking of Ignorance*, edited by L. Schiebinger and R. N. Proctor, 1 - 34. Stanford: Stanford University Press.

Sauermann, H., and C. Franzoni. 2015. "Crowd Science User Contribution Patterns and their Implications." *Proceedings of the National Academy of Sciences* 112 (3):679-684.

Tesh, S. N. 2000. *Uncertain Hazards: Environmental Activists and Scientific Proof*. Ithaca: Cornell University Press.