Citizen Science In The Digital Age: Connecting ‘Unapplied Capacities’ With ‘Unmet Needs’

Jack Fotheringham – University of Saskatchewan

Graeme Jobe – University of Saskatchewan

CDO Partnership Conference Paper for the Student Research Presentation Session

May 1, 2017

Citizen Science In The Digital Age: Connecting 'Unapplied Capacities' With 'Unmet Needs'

Abstract

Citizen Science (CS) projects are playing an increasingly important role in both the advancement of scientific knowledge and in community engagement for scientific research organizations. With the advent of Web 2.0—and particularly since the ubiquitous consumer adoption of smartphone technology—the past 25 years have seen a great increase in the number of citizen science projects, with the majority of CS projects being web-based and globally accessible. Beyond accessibility for participation, the Internet has provided platforms--most notably Zooniverse--that make it easy for those who lack the capacity to produce or analyze large datasets to achieve their goals through volunteer participation.

In the context of CS, digital technology has been used to overcome financial, geographic, and temporal barriers to research. CS projects provide a way for the global community to participate in research that is specific to a local context, increasing local impact while strengthening global interconnectivity. Digital technology provides everyday people with a means to interface with scientific research, and to aid researchers in collecting and analyzing data in fields from ecology to astrology. All projects contribute to the actionable, analyzed open data as a public good—producing an output that advances public and private research and enriches the knowledge economy.

With the increase in open access data produced by public research institutions such as universities and NGOs and concomitant growth of crowdsourcing data collection and analysis in the private sector, examining CS projects may produce valuable insights into how human capital functions within nascent aspects of the digital economy. This report provides an overview of characteristics of existing CS projects along with some suggestions as to how insights gleaned from their study can be utilized from a public policy perspective.

Keywords: citizen science; innovation; digital technology; innovation policy

1. An Overview of Citizen Science

Citizen Science (CS) projects, also referred to as 'public participation in scientific research,' have been of increasing importance to scientific advancement in recent years. While the data gleaned from CS projects is well verified and generally open access, only a small proportion of it (approximately 12 per cent in biodiversity-related projects, for instance) reaches publication in a scientific journal.[1] Despite their comparative lack of representation in peer reviewed scientific journals, CS projects have given rise to some notable scientific discoveries on the part of citizen scientists themselves, including the discovery of a new type of astronomical object by a Dutch school teacher in 2007.[2] In addition to advancing scientific research itself, CS projects are shown to have other positive functions. For instance, CS projects have been shown to create increases in scientific literacy and positive attitudinal changes towards science and learning in participants.[3] Furthermore, CS projects that deal with biodiversity and ecological topics (the majority of CS projects) could even increase public participation and effectiveness of policy making in conservation.[4]

The majority of studies in the nascent Citizen Science literature focus on characteristic aspects such as the reliability of data they produce, their advantages over conventional research project structures, and their effect on participants' attitudes and literacy. Instead this report seeks to build upon previous work by Fotheringham, which surveyed a of wide variety of CS projects to

---

[1] E.J. Theobald et al., "Global Change and Local Solutions: Tapping the Unrealized Potential of Citizen Science for Biodiversity Research," *Biological Conservation* 181 (January 2015): 236–44, doi:10.1016/j.biocon.2014.10.021.
[2] Dale R. Wright et al., "Understanding the Motivations and Satisfactions of Volunteers to Improve the Effectiveness of Citizen Science Programs," *Society & Natural Resources* 28, no. 9 (September 2, 2015): 1013–29, doi:10.1080/08941920.2015.1054976.
[3] A. W. Crall et al., "The Impacts of an Invasive Species Citizen Science Training Program on Participant Attitudes, Behavior, and Science Literacy," *Public Understanding of Science* 22, no. 6 (August 1, 2013): 745–64, doi:10.1177/0963662511434894.
[4] Ayesha I.T. Tulloch et al., "Realising the Full Potential of Citizen Science Monitoring Programs," *Biological Conservation* 165 (September 2013): 128–38, doi:10.1016/j.biocon.2013.05.025.

create a conceptual framework for their categorization based on the function of participants.[5] Where this work strove toward greater conceptual clarity within the CS space with a view to the development of policies that would better support CS initiatives, our current paper is an attempt to tease out how these insights relate to the broader topics central to the Creating Digital Opportunity (CDO) project. Specifically, we are interested in creating a preliminary idea of how CS is situated within the current digital landscape, and what we might learn from it when studying other areas of digital innovation and crowdsourcing.

2. Citizen Science in the Digital Economy

As discussed above, the intention of this paper is to develop an account of how CS can contribute to the ongoing discussion around innovation in the digital economy. It must be disclaimed at the outset that this account is preliminary—both as it relates to the literature and in terms of the available evidence for emergent trends and possibilities. Our aim is merely to advance clear argumentation demonstrating the possibility—if not probability—of a range of diverse outcomes and their implications. This will be achieved first through bringing greater conceptual clarity to differentiated modes of CS participation, and second through illustrative examples corresponding to each category, appearing, respectively, in the Conceptual Framework and Analysis sections to follow. However, this section will first set out a multitude of potential connections between our existing understanding of CS and the expansive digital economy—both within and across the public and private spheres. As a first step toward exploring these connections through scholarly inquiry, we suggest that modelling how CS initiatives relate to participants can aid in understanding the adoption of similar strategies in the private sector.

[5] Jack Fotheringham, "Survey Citizen Science Initiatives: Toward a Participant-Focussed Conceptual Framework," *(unpublished)*, August 2016, 7.

Similar to all innovative activities that characterize the digital economy, CS would not have been possible without the widespread adoption of smartphones and Web 2.0 platforms, which have enabled instant and ongoing global interconnectedness.[6] Though we will later demonstrate how CS is characterized by unique modes of participant activity and interaction, it can also be characterized by a simple notion that applies equally to similar non-CS activities such as end-user innovation, mass collaboration, and crowdsourcing.[7] In the most fundamental sense, we consider all such activities to be means of connecting 'unmet needs' with 'applied capacities.' In and of the digital economy, this joining is maintained on an ongoing basis, in real time, and across an interconnected globe. Further, connections need not operate unidirectionally in discrete phases from 'needs' to 'capacities' or vice versa, and often operate bidirectionally and on a continual basis.

The Zooniverse platform exemplifies this in the CS space, enabling willing participants to seek out exciting initiatives to which they can apply their capacities, whether they are highly skilled or simply interested in tasks requiring a human eye or basic decision-making capabilities. At the same time, researchers, conservationists, or users with raw, unanalyzed datasets can leverage Zooniverse to transcend temporal and spatial boundaries in advertising to and connecting with willing participants. In addition to easing these barriers, CS projects find additional benefits not often present in more traditional research and data-gathering domains. Such benefits include an increased scale and scope of available participants, reduced transaction and labour costs, and an increased public interest in the subject area.[8]

---

[6] McKinsey Global Institute, "Disruptive Technologies: Advances That Will Transform Life, Business and the Global Economy" (Boston: McKinsey & Co., 2013).
[7] Stephan Thomke and Eric von Hippel, "Innovation: A New Way to Create Value," *Harvard Business Review* 80, no. 4 (April 2002): 74–81.
[8] Crall et al., "The Impacts of an Invasive Species Citizen Science Training Program on Participant Attitudes, Behavior, and Science Literacy."

In theory, the same benefits should also be applicable to private sector activities, and we will introduce a number of illustrative examples to demonstrate this later on. However, the differing nature of private sector objectives and incentives locates the ‘needs-capacities interface’ across a very different range of activities. For instance, end-user innovation of product design provides a fascinating example of how digitally-enabled interconnection of ‘unmet needs’ and ‘unapplied capacities’ can be achieved.[9] In this relationship, the firm and customer each occupy both roles on an ongoing basis, with the aggregate of end-users providing constant feedback based on their experiences with products. This feedback then constitutes an input to improved product design on the firm side. In other words, when firms and customers share a common interest in product improvement, customer needs are transformed into a capacity to guide design activities. Further, two-way information sharing could enable new technologies and products to respond to customer needs throughout each stage of development while simultaneously conditioning customer demands through real-time updates on the emergence of new technological advances. Indeed, this sort of bi-directional relationship may someday render the distinction between ‘supply-push’ and ‘demand-pull’ products an anachronism, though we have yet to find a real-world firm that embodies this.

In contrast, CS initiatives are generally not profit-driven, existing instead to provide and enrich public goods such as ecological conservation, public health, and general advancement of the knowledge frontier. Goods that are both nonrivalrous and nonexcludable will be undersupplied in the free market, and thus can only be supplied at socially optimal levels by the public sector.[10] Indeed, fields common to CS such as astronomy and phenology tend to be associated with low commercial exploitability, and to this extent, may be unlikely to attract much interest from the private sector. However, it is unclear how exactly this applies to open datasets, the use of which is highly contextual

[9] Thomke and von Hippel, “Innovation: A New Way to Create Value.”
[10] Hal R. Varian, *Microeconomic Analysis*, 3rd ed (New York: Norton, 1992).

and multifaceted. Exactly how rivalrousness and excludability apply to digital innovation remains unclear, in part, because many of the most fruitful uses of data could become rivalrous and excludable due to technological barriers. For instance, big data analysis requires vast amounts of computing power and storage necessary to draw insights from massive datasets.[11] As such, it is entirely conceivable that the large, open datasets spawned through CS-initiatives could be leveraged differentially by private actors, as alluded to earlier.

What remains in this paper explores the above themes at a more concrete level, using real cases that illustrate the joining of unmet needs with unapplied capacities both in the CS space and private sector. However, before this departure from conceptual to concrete, the next section will introduce a conceptual framework based on a previous survey of CS projects by Fotheringham, which focusses on the role of participants.[12]

3. Method

While there exist many studies of individual CS projects which examine the quality of data produced, objectives achieved, and organizational structure, there are few that provide a framework for sorting these projects.[13] Of the studies that do categorize CS projects, the majority focus on differentiating types of relationships between citizen scientists and project organizers or attitudinal impacts on CS participants.[14] In contrast, this report seeks to categorize CS projects based on the

[11] Viktor Mayer-Schonberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Boston, Massachusetts: Houghton Mifflin Harcourt, 2013).
[12] Fotheringham, "Survey Citizen Science Initiatives: Toward a Participant-Focussed Conceptual Framework."
[13] Tulloch et al., "Realising the Full Potential of Citizen Science Monitoring Programs"; Duncan C. McKinley et al., "Citizen Science Can Improve Conservation Science, Natural Resource Management, and Environmental Protection," *Biological Conservation* 208 (April 2017): 15–28, doi:10.1016/j.biocon.2016.05.015.
[14] C. Aaron Price and Hee-Sun Lee, "Changes in Participants' Scientific Attitudes and Epistemological Beliefs during an Astronomical Citizen Science Project: CITIZEN SCIENCE

function of CS researchers within the project. Rather than pre-establishing these categories and then populating them with examples, this study examined approximately 30 different CS projects, and inductively created categories to highlight such distinctions. The CS projects surveyed were sufficiently different to establish four categories: Active Data Collection, Passive Data Collection, Skilled Data Analysis, and Unskilled Data Analysis.

4. Framework and Results

Active Data Collection

As the name suggests, participants in Active Data Collection projects play an active role in collecting data to be used in the study. These projects usually require participants to have some knowledge of the field they are working in, and generally provide them with this information as a part of the project. Participants do not analyze the data, but submit it (often electronically) to the researchers. In this model, 'participatory monitoring' projects, such as migratory bird counts and volunteer water sampling programs, are included in this category.

Passive Data Collection

Participants in Passive Data Collection projects do not make regular, pointed efforts to collect data, but rather consent to using some kind of sensor to monitor some aspect of their environment. These types of projects are sometimes considered to be distinct from Citizen Science because they often do not actively involve their participants.

LITERACY," *Journal of Research in Science Teaching* 50, no. 7 (September 2013): 773–801, doi:10.1002/tea.21090.

Skilled Data Analysis

Participants in Skilled Data Analysis projects analyze data provided to them by the research organization running the project. This analysis requires a substantial degree of relevant knowledge in order to contribute. Some Skilled Data Analysis projects provide participants with training modules in order to gain this knowledge, while others require that participants pass a test of their knowledge in order to participate. Both of these practices ensure that participants' contributions are accurate.

Unskilled Data Analysis

Participants in Unskilled Data Analysis projects do not require more than a layperson's knowledge of the data they are analyzing. For instance, such analysis can be the identification of simple colours or objects in photographs, or the transcription of handwritten documents. These projects often require analyses to be verified by multiple participants to ensure accuracy.

Category	Participants collect data	Participants analyze data	Researchers collect data about participants	Requisite level of knowledge for participation
A.D.C.	Yes	No	No	Sometimes
P.D.C.	Yes	No	No	Sometimes
S.D.A.	No	Yes	Sometimes	Yes
U.D.A.	No	Yes	Sometimes	No

4.1

This study examined the above defining characteristics of 145 Citizen Science projects, and sorted each into the appropriate category. The region of origin and accessibility, date of origin, and topic of study were also documented in the interest of finding trends in where and when different types of CS projects have been created, along with their purpose. Additionally, we have included in the Analysis section a brief overview of specific CS projects that exemplify the characteristics of their respective categories and have produced significant findings in their fields of research.

Category	**Number**
Active Data Collection	104
Passive Data Collection	7
Skilled Data Analysis	11
Unskilled Data Analysis	23

Active Data Collection

As the shown in the above chart, the vast majority (104 of 145) of projects categorized fell into the Active Data Collection category. The oldest and most well established types of CS projects such as migratory bird counts fall under this category, which helps to explain their prominence. The earliest of these projects include the Cooperative Observer Program, dating back to 1890, and the Audubon Christmas Bird Count, launched in 1900. As these projects generally do not require special knowledge for participation, nor intensive time commitments, they are easily accessible to participants.

Of these 104 CS projects, 92 studied ecology, eight studied astronomy, and four studied health sciences. Not only do ecological projects make up 88 per cent of the projects in this category - this subset also comprises approximately 63 per cent of all of the projects surveyed. It is likely that ecological projects in the Active Data Collection category are so numerous because they often have major advantages over conventional research in this field. By nature, ecological surveys require large datasets to ensure the accuracy of their findings. Active Data Collection CS projects give researchers access to datasets that are much larger, both geographically and temporally, than they would otherwise have. For instance, the National Audubon Society's Christmas Bird Count has been conducting counts of migratory bird populations across North America since the year 1900. Without volunteer participation, which has sustained the Christmas Bird Count throughout its existence, the project could not possibly be so expansive or long-running.

Passive Data Collection

Passive Data Collection comprised only seven of 145 CS projects, with two in astronomy, three in ecology, and two in health sciences. Notably, all seven projects were international in scope and origin, and all were based online. As with Active Data Collection, these projects provided researchers with access to large datasets acquired over the course of years by individuals from around the world.

Skilled Data Analysis

Of the eleven projects in this category, six studied astronomy and five studied ecology, which reflects the general preponderance of these topics in citizen science on the whole. Some of the projects relied on participants' pre-existing knowledge of the subject matter, and all provided a web-based tutorial to teach participants how to analyze the data. The most common function of the participants was to identify and categorize objects in photos, ranging from animals to galaxies.

Notably, the projects in this category were all created in 2011 or later, with the exception of one (Stardust@Home) created in 2006. These projects represent a significant advancement in the character of CS, and the role of citizen scientists in Skilled Data Analysis is much more a conscious part of the scientific method than a means for data collection. Of the eleven projects, six were based internationally, and all were web-based and accessible worldwide. This category may well represent the future of CS, as it educates the public while benefiting from their involvement.

Unskilled Data Analysis

Accounting for 23 of 145 CS projects surveyed, Unskilled Data Analysis was the second largest of the four categories created. All 23 projects were web-based and accessible around the world, with 12 based internationally. This closely mirrors the characteristics of projects in the Skilled Data Analysis category, with the notable absence of in-depth tutorials and mandatory skill testing prior to participation. Furthermore, 20 of the 23 projects were created in 2012 or later, indicating a trend within CS toward this type of project in general. The projects in this category were also by far the most diverse in topic, with three in astronomy, nine in ecology, three in health sciences, six in history, one in linguistics, and one in physics. This suggests potential for projects of this format to be suitable for a wide variety of applications.

5. Analysis

In order to build on the conceptual framework advanced in the previous section to analyze corollaries between functions of participants in CS projects and equivalent functions of user-consumers in analogous private enterprises, this section examines case studies from each of the four established categories, with the notable addition of 'platforms'. It is our aim that with features of specific CS projects in mind, analogies and disanalogies with existing and prospective private enterprises will be more readily apparent. After establishing these benchmarks for comparison, we proceed to describe and compare existing corollaries in the private sector and explore the potential for corollaries that have not yet been established.

5.1 CS Project Case Studies

Active Data Collection

eButterfly (2011) Canada & USA

University of Ottawa, University of Arizona, Espace Pour La Vie (Montreal), Vermont Center for Ecostudies

Participants use web-based software to track the presence or absence of various species of butterflies in their area by completing online checklists and uploading photos of the species they encountered. Species identification is verified by volunteer regional experts over the web before the data is accepted, and unusual reports are confirmed by volunteer taxonomic experts. eButterfly is currently developing additional web-based species identification training tools for use by its participants to further increase accuracy. eButterfly also integrates data with local, regional, and national organizations that conduct similar research to create a much larger dataset, and is planning to release its data to larger biodiversity organizations. This data is valuable for research in conservation, phenology, biology, and agriculture. As with most CS projects, all data is open access.[15]

Passive Data Collection

Safecast (2011) Worldwide

Shuttleworth Foundation

[15] eButterfly, "eButterfly Latest News," *E-Butterfly.ord*, n.d., http://www.e-butterfly.org/.

Safecast was founded in 2011 in response to the earthquake and tsunami which affected Japan in 2011 and caused the meltdown of the Fukushima Daiichi Nuclear Power Plant. Safecast gathers data on environmental radiation and other air pollutants through the observations of citizen scientists around the world. In addition to providing the most comprehensive information on environmental radiation available, Safecast enables participants to monitor the air quality and radiation levels in real time in their own homes.

Most interestingly, Safecast does not manufacture the hardware that participants use to collect data. Instead, developers at Safecast create open source designs for sensor modules, hosted on Github, which participants can use to inexpensively build their own modules at home. Safecast 'maintains the largest open dataset of background radiation measurements ever collected' and has collected over 50 million readings to date. All of the data collected is published in the public domain and can be used without licensing restrictions.[16]

Skilled Data Analysis

Agent Exoplanet (2011) USA

Los Cumbres Observatory Global Telescope Network

Citizen scientists can help to discover and classify exoplanets by analyzing images taken in California, Hawaii, and Australia, by the Los Cumbres Observatory Global Telescope Network, as a part of the CS project 'Agent Exoplanet.' Citizen scientists are trained online to use web-based software to detect the presence of exoplanets, and to analyze and classify 'lightcurves', which provide researchers with valuable information with which to estimate the size, orbital direction, velocity, and indirectly the chemical composition of its atmosphere. This is known as the 'transit

[16] Safecast, "Safecast - Open Environmental Data for Everyone," *Blog.safecast.org*, n.d., blog.safecast.org.

method' for finding exoplanets. Findings are verified by at least three other participants to ensure accuracy.[17]

Unskilled Data Analysis

Old Weather (2012) Worldwide

National Archives and Records Administration, The MET Office, NOAA, National Maritime Museum, New Bedford Whaling Museum, Providence Public Library, University of Oxford, et al.

Participants transcribe digital images of handwritten ship logs from the 19th and 20th centuries, while collecting information on weather conditions and sea ice conditions observed by ships' crews in the past. Participants also discover and catalogue information of historical value about ships' routes and the events that occurred on various voyages. Computers are incapable of accurately transcribing human handwriting, especially from this time period, and so citizen scientists form an invaluable part of the data collection process. The resulting larger datasets make reanalysis and simulations of climate data more accurate.[18]

CS Platform

Zooniverse (2009) Based in US and UK, Accessible Worldwide

University of Nottingham, ETH Zürich, University of Oxford, Adler Planetarium, University of Minnesota, Johns Hopkins University, Vizzuality, ASIAA, University of Portsmouth

Zooniverse in a web portal that serves as "the world's largest and most popular platform for people-powered research." Zooniverse is owned by the Citizen Science Alliance, which seeks to promote CS projects—all of which fall under the Skilled or Unskilled Data Analysis categories—for

[17] Agent Exoplanet, "Agent Exoplanet Mission Brief," *Agentexoplanet*, n.d., https://lco.global/agentexoplanet/.
[18] Old Weather, "Old Weather," n.d., https://www.oldweather.org/.

their ability to analyze large data sets, yield qualitative estimates of error in analysis, create 'training sets' for the development of machine learning algorithms, produce serendipitous discoveries, and educate participants. In addition to user accessibility, the consolidation of many CS projects provides benefits to researchers in the form of shared and purpose-built infrastructure for hosting projects. While many of the projects hosted by Zooniverse are Citizen Science Alliance initiatives, Zooniverse also features a built-in 'Project Builder' that allows any researcher to create a CS project to be hosted by Zooniverse using a proven and convenient framework.[19]

5.2 Existing And Prospective Corollary Examples From The Private Sector

Active Data Collection

Poimapper (2009) Based In Finland, Accessible Worldwide

Pajat Solutions Ltd.

Poimapper is a product developed by Finnish software firm Pajat Solutions Ltd that enables organizations to obtain data from their fieldworkers in real time via mobile devices. Client organizations can set parameters for what kind of data they wish to collect, be it photos, location information, timestamps, or user input. This data is then aggregated in a cloud server, and can be analyzed by the client organization and shared between fieldworkers in real time. Poimapper has been used for purposes as diverse as tracking industrial gas shipments, building inspection, and recording progress and outcomes in international development projects. Although Poimapper provides a service to other organizations rather than conducting their own projects, the role of in-field data collectors is identical to that in CS Active Data Collection projects. Interestingly, Poimapper is an example of active data collection within a closed—and often profit-oriented—

[19] University of Oxford, "A-Z of Oxford - Z Is for Zooniverse," *Ox.ac.uk*, n.d., http://www.ox.ac.uk/content/z-zooniverse.

system. It is important to note, however, that while Poimapper's clients do take advantage of real time data submission, streamlined data entry, and a wider array of potential inputs (GPS coordinates, photographs, etc.), they do not share with CS Active Data Collection projects the advantage of massively temporally and geographically dispersed data collection as data is collected by employees, not volunteers.[20]

Passive Data Collection

Google Traffic (2007) Worldwide

Google Maps

Perhaps the most promising conduit for passive data collection in the present and immediate future is the ubiquity of smartphones, as evidenced by the success of Google Traffic. Google Traffic was created by Google after it acquired ZipDash, a private tech firm that focused on providing real-time traffic analysis, in 2004. Google Traffic is a feature of Google Maps that collects information such as GPS coordinates, velocity, and frequency of stops from motorists' smartphones, and uses the aggregate data to provide traffic information that is updated in real time to users of Google Maps. This is one of the most accurate ways of monitoring traffic, and yet requires no additional equipment or infrastructure (e.g., cameras monitoring roadways) beyond individuals consenting to the transmission of their location data to Google Maps on a continual basis.[21]

Skilled Data Analysis

---

[20] Poimapper, "How Poimapper Works," *Poimapper.com*, n.d., http://www.poimapper.com/start-using-poimapper/.

[21] Tim Stenovec, "Google Has Gotten Incredibly Good at Predicting Traffic — Here's How," *Businessinsider.com*, December 18, 2015, http://www.businessinsider.com/how-google-maps-knows-about-traffic-2015-11.

Unlike the other categories, we were unable to find an adequate corollary for Skilled Data Analysis in the private for-profit sector. However, there were various enterprises that encapsulated some features of projects in this category. For instance, Apple Inc. offers a service called "Apple Support Communities," in which owners of Apple products who are having trouble with a device may seek or provides helpful responses to troubleshoot the issue. In this model, participants are highly knowledgeable about the use of Apple products, but comparatively little or no data is produced or analyzed. Similarly, end-user innovation projects like the ones described in von Hippel appear to share features with the category while not being wholly aligned.[22] Perhaps most appropriate prospective form of private enterprise resembling a Skilled Data Analysis CS project has been outlined in Willett, Heer, and Agrawala.[23] In this prospective case, Willet, Heer, and Agrawala make a strong case for crowdsourcing the analysis of trends in data itself, showing that it is in some instances much more effective than analysis by an individual or small group.

Unskilled Data Analysis

reCAPTCHA (2007) Worldwide

Google Inc.

Although its current format is different, reCAPTCHA used the original CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) system to protect websites from 'bots' which could do anything from 'scalp' event tickets to collect private e-mail addresses for spam e-mail. This specific system was originally developed by a group of computer scientists at Carnegie Mellon in Pittsburgh, US, but was purchased by Google in 2009. Originally, reCAPTCHA worked by having users input their interpretation of a selection of distorted text that

[22] Thomke and von Hippel, "Innovation: A New Way to Create Value."
[23] Wesley Willett, Jeffrey Heer, and Maneesh Agrawala, "Strategies for Crowdsourcing Social Data Analysis" (ACM Press, 2012), 227, doi:10.1145/2207676.2207709.

computers are unlikely to be able to read. While providing an incentive for website administrators to use the free service by protecting them from fraud and spam, this formulation of reCAPTCHA enabled Google to digitally transcribe the entire archive of the New York Times and every book on Google Books, along with creating an enormous training data set for machine learning algorithms to read text. Through a relatively limited imposition on website-users—having to input a word that would appear on a screen to access the site—Google managed to crowdsource the digitization of over 13 million articles and the creation of a valuable data set.[24]

Platform

Amazon Mechanical Turk

Amazon.com, Inc.

Similar in many ways to Zooniverse, Amazon Mechanical Turk (MTurk) is a platform that enables parties with unmet needs to find those with the capacities to meet them. MTurk functions by enabling employers (known as 'requesters') to post jobs (known as Human Intelligence Tasks) to find employees ('providers') to carry them out. Human Intelligence Tasks are tasks that cannot be performed by a computer, and range in their requirements of skill. As is the same in the case of Zooniverse, MTurk provides the option of requiring multiple confirmations by other providers in order for a result to be accepted as true. In addition to providing an alternate market for online short-term contract labour, MTurk also presents an opportunity for requesters to collect large training data sets for machine learning by having providers analyze data like photos.[25]

[24] Google, "reCAPTCHA," *Google.com*, n.d., https://www.google.com/recaptcha/intro/invisible.html.
[25] mturk.com, *Www.mturk.com*, n.d., https://www.mturk.com/mturk/help?helpPage=overview#what_is.

6. Discussion And Policy Implications

This paper began by exploring how CS relates to the broader digital economy by abstracting to the level of 'unmet needs' and 'unapplied capacities' brought together through digital innovation. We subsequently introduced a conceptual framework used to differentiate modes by which participants engage in CS initiatives, with a view to illuminating a pathway for private sector adoption of similar models—both presently realized and prospective. The paper then populated this framework—first with cases located in the CS space, followed by analogues emerging in the private sector. This leaves us now at the critical question of 'so what'?

To this challenge, we submit that that the conceptual framework developed in Fotheringham proved eminently applicable to a number of private-sector initiatives we were able to identify with relative ease.[26] Initially developed inductively through the process of surveying an uncategorized assortment of CS cases, our framework was then utilized to deductively characterize a handful of cognates that featured private-sector customers engaged in comparable modes of participation. That conceptually similar activities to those in CS are emerging in the private domain must be distinguished from the claim that such activities resulted from the diffusion of innovations from the CS space. Though it is possible this has occurred or is likely to in the future, it is more likely that the emergence of new avenues for the joining of 'unmet needs' with 'unapplied capacities' has occurred simultaneously as a product of the same pervasive digital innovations. Perhaps a more relevant question centres around how differences—rather than similarities—between these domains relates to governance.

The challenge in governing digital change is characterized by the necessity to find balance in the creative tension between open innovation (i.e. open access to public knowledge) and market

[26] Fotheringham, "Survey Citizen Science Initiatives: Toward a Participant-Focussed Conceptual Framework."

incentives (i.e. value creation through commercialization of knowledge).[27] Much in the same way intellectual property (IP) facilitates public access to new knowledge in exchange for a temporary monopoly over the value it generates, a framework for the governance of digital technology should allow for public and private benefits to be mutually reinforcing. Policies must ensure that current innovations spill over into further discoveries while continuing to attract investment and add economic value. CS can contribute to this end, both in and of itself and by virtue of insights it can provide into analogous market activity spanning outside its conceptual bounds.

Ultimately, CS and its analogues in the private sector present new avenues through which human activity can add value to public good initiatives at little or no cost to participants, while often enriching their lives through learning and meaningful experiences. Adopting lessons and principles from CS to private economic activities involves understanding CS models of participation to better leverage human capital. From a social policy viewpoint, this represents a use of digital technology that increases the value of human labour—and potentially demand for it—by creating new job opportunities with very little barrier to entry. On the other hand, increased worker productivity could decrease the quantity of labour demanded in some positions that are made more efficient by CS practices. Or at the more drastic end of the spectrum, adoption of such models could increase the rate at which computers learn from human behaviour by producing machine learning algorithms that could render these forms human labour redundant at some point in the future.

6.1 Lines For Further Inquiry

As this paper generally explores the connections between CS, advancements in the digital space, and public policy, there remains plenty of work to be done in identifying potentially

[27] Peter W. B. Phillips, *Governing Transformative Technological Innovation: Who's in Charge* (Northampton, Massachusetts: Edward Elgar Publishing Ltd., 2007).

problematic, interesting, or advantageous developments in the employment of digital technology to maximize human capacities. We here outline three lines of further inquiry, without discounting the plenty of others that exist. First, there is the question of whether appropriation of data-driven products and open access to datasets generated by CS initiatives could differentially benefit commercial interests. As much of the data produced by CS projects are open access, private firms may use this data—which has already been analyzed by participants—as data training sets to develop machine learning. Although data for training sets is not impossible to find elsewhere, it can be expensive to obtain as it requires human intelligence to develop. Whether this will be a positive feature in that open access CS data sets will act as a communal training data set remains to be seen.

Second, end user innovation of the kind described by von Hippel may become an increasingly common component of the product design process as barriers to communication between end users and designers are reduced by digital technology.[28] While this sort of innovation already has a considerable presence in computer software design and the like, we are interested to see whether private firms that produce specialized products with no inbuilt feedback mechanisms—such as musical instruments, firearms, sewing machines—may turn to digital technology to incorporate user innovation. Third, independently-organized CS initiatives could either disrupt or become further integrated within existing systems that provide public goods such as research, education, or public safety. While to our knowledge CS projects to date have only supplemented existing provision of public goods, it is possible that CS or CS-like projects could take on a greater role in or even displace conventional components of research and education.

7. Conclusion

This paper has taken steps toward understanding developments in digital technology through

[28] Thomke and von Hippel, "Innovation: A New Way to Create Value."

a lens focused on the role of human capital in innovative technological endeavours. While this study is by no means exhaustive, it may serve to guide future inquiry in this field. As the digital landscape changes with new advances, understanding the role of human participants within enterprises centred on the collection and analysis of data will continue to inform policy decisions. While there is a role for policy makers to play in ensuring that advances in these technologies do not come at the expense of the public good, they must also strike a balance with ensuring the potential for success in the private sector.

Bibliography

Agent Exoplanet. "Agent Exoplanet Mission Brief." *Agentexoplanet*, n.d. https://lco.global/agentexoplanet/.

Crall, A. W., R. Jordan, K. Holfelder, G. J. Newman, J. Graham, and D. M. Waller. "The Impacts of an Invasive Species Citizen Science Training Program on Participant Attitudes, Behavior, and Science Literacy." *Public Understanding of Science* 22, no. 6 (August 1, 2013): 745–64. doi:10.1177/0963662511434894.

eButterfly. "eButterfly Latest News." *E-Butterfly.ord*, n.d. http://www.e-butterfly.org/.

Fotheringham, Jack. "Survey Citizen Science Initiatives: Toward a Participant-Focussed Conceptual Framework." *(unpublished)*, August 2016, 7.

Google. "reCAPTCHA." *Google.com*, n.d. https://www.google.com/recaptcha/intro/invisible.html.

Mayer-Schonberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, Massachusetts: Houghton Mifflin Harcourt, 2013.

McKinley, Duncan C., Abe J. Miller-Rushing, Heidi L. Ballard, Rick Bonney, Hutch Brown, Susan C. Cook-Patton, Daniel M. Evans, et al. "Citizen Science Can Improve Conservation Science, Natural Resource Management, and Environmental Protection." *Biological Conservation* 208 (April 2017): 15–28. doi:10.1016/j.biocon.2016.05.015.

McKinsey Global Institute. "Disruptive Technologies: Advances That Will Transform Life, Business and the Global Economy." Boston: McKinsey & Co., 2013.

mturk.com. *Www.mturk.com*, n.d. https://www.mturk.com/mturk/help?helpPage=overview#what_is.

Old Weather. "Old Weather," n.d. https://www.oldweather.org/.

Phillips, Peter W. B. *Governing Transformative Technological Innovation: Who's in Charge*. Northampton, Massachusetts: Edward Elgar Publishing Ltd., 2007.

Poimapper. "How Poimapper Works." *Poimapper.com*, n.d. http://www.poimapper.com/start-using-poimapper/.

Price, C. Aaron, and Hee-Sun Lee. "Changes in Participants' Scientific Attitudes and Epistemological Beliefs during an Astronomical Citizen Science Project: CITIZEN SCIENCE LITERACY." *Journal of Research in Science Teaching* 50, no. 7 (September 2013): 773–801. doi:10.1002/tea.21090.

Safecast. "Safecast - Open Environmental Data for Everyone." *Blog.safecast.org*, n.d. blog.safecast.org.

Stenovec, Tim. "Google Has Gotten Incredibly Good at Predicting Traffic — Here's How." *Businessinsider.com*, December 18, 2015. http://www.businessinsider.com/how-google-maps-knows-about-traffic-2015-11.

Theobald, E.J., A.K. Ettinger, H.K. Burgess, L.B. DeBey, N.R. Schmidt, H.E. Froehlich, C. Wagner, et al. "Global Change and Local Solutions: Tapping the Unrealized Potential of Citizen Science for Biodiversity Research." *Biological Conservation* 181 (January 2015): 236–44. doi:10.1016/j.biocon.2014.10.021.

Thomke, Stephan, and Eric von Hippel. "Innovation: A New Way to Create Value." *Harvard Business Review* 80, no. 4 (April 2002): 74–81.

Tulloch, Ayesha I.T., Hugh P. Possingham, Liana N. Joseph, Judit Szabo, and Tara G. Martin. "Realising the Full Potential of Citizen Science Monitoring Programs." *Biological Conservation* 165 (September 2013): 128–38. doi:10.1016/j.biocon.2013.05.025.

University of Oxford. "A-Z of Oxford - Z Is for Zooniverse." *Ox.ac.uk*, n.d. http://www.ox.ac.uk/content/z-zooniverse.

Varian, Hal R. *Microeconomic Analysis*. 3rd ed. New York: Norton, 1992.

Willett, Wesley, Jeffrey Heer, and Maneesh Agrawala. “Strategies for Crowdsourcing Social Data Analysis,” 227. ACM Press, 2012. doi:10.1145/2207676.2207709.
Wright, Dale R., Les G. Underhill, Matt Keene, and Andrew T. Knight. “Understanding the Motivations and Satisfactions of Volunteers to Improve the Effectiveness of Citizen Science Programs.” *Society & Natural Resources* 28, no. 9 (September 2, 2015): 1013–29. doi:10.1080/08941920.2015.1054976.