Human Computation (2015) 2:1:19-44

ISSN: 2330-8001, DOI: 10.15346/hc.v2i1.3

# Privacy in Participatory Research: Advancing Policy to support Human Computation

ANNE E. BOWSER, WOODROW WILSON INTERNATIONAL CENTER FOR SCHOLARS

ANDREA WIGGINS, UNIVERSITY OF MARYLAND

## ABSTRACT

Citizen science and participatory sensing are two models of human computation in which participant privacy is a key concern. Technological safeguards are important but partial solutions; a full and accurate description of policies explaining privacy practices must also be present so volunteers can make informed decisions regarding participation. Our study surveyed the policies of 30 participatory research projects to establish how privacy-related policies were presented, and how they aligned with actual practices. This paper contributes a description of the privacy-related elements of policies evident in these projects. We found that while the majority of projects demonstrated some understanding of the need for policies, many hosted incomplete policies or inaccurately described their practices. We discuss the implications for project management, design, and research or operational policy, both for projects in citizen science and participatory sensing, and for the larger field of human computation. We conclude by proposing a set of Ethical Practices for Participatory Research Design as guidelines to inform the development of policies and the design of technologies supporting participatory research.

## 1. INTRODUCTION

Facebook has made the news for its privacy practices a number of times. In 2014, outlets including Forbes,[1] WIRED,[2] and The New Scientist[3] pounced on a study of emotional contagion where Facebook data scientists manipulated the news feeds of over 600,000 users to demonstrate how positive and negative affect spreads between “friends.” Some critics argued that users’ privacy

---

[1] http://www.forbes.com/sites/kashmirhill/2014/05/22/facebook-wants-to-listen-in-on-what-youre-doing/

[2] http://www.wired.com/2014/07/business-facebook-feelings/

[3] http://www.newscientist.com/article/dn25815-dont-fear-facebooks-emotion-manipulation-experiment.html#.U-32Pa1dVfR

and well-being were unreasonably threatened. But many also noted that Facebook's practices were in line with their data use policy, which clearly states that personal information will be used "for internal operations, including troubleshooting, data analysis, testing, research and service improvement."

While Facebook can afford a team of legal professionals to codify its data use policies, not every researcher enjoys such support. For example, human computation is a growing paradigm where human information processing augments computational systems to meet real world goals (Michelucci, 2013). This encompasses many forms of research, including citizen science and participatory sensing, where public volunteers conduct research through data collection, analysis, interpretation, and/or application (e.g., through research-driven advocacy). These projects, which we refer to as *participatory research projects*, are typically led by university researchers, government agencies, nonprofit organizations, and community groups.

As the strategy of turning to volunteers might suggest, these groups may be short on resources, such as the expertise to develop policies and participant protections. They might also fail to recognize the value of such policies, especially when development of a novel technology is the focal point of research. As the director of one participatory sensing lab asserted, "there would *be* no data without first having a system" (as observed by Shilton, 2012, p.8). Finally, researchers and volunteers engaged in participatory research may lack access to ethics evaluation tools, including Institutional Review Boards (IRBs), that their colleagues conducting more standardized forms of research enjoy.

Yet, as illustrated by the Facebook example, some research methodologies and supporting technologies may protect (or exploit) privacy in a way that conflicts with user expectations, or the expectations of society as a whole. While this is an issue worth addressing in all contexts, it is particularly salient in participatory research because:

i. These projects often collect sensitive data, such as precise GPS locations, but it is not always intuitively clear to volunteers why this information may be sensitive. For example, a volunteer submitting an observation of a bird in real time may consider the data to be about the bird, rather than about the human volunteer (Bowser, Wiggins, Shanley, Henderson, & Preece, 2013). Location-based data can also be context-dependent; the same participant may have no qualms sharing her location during working hours, but hesitate to share her location when engaged in leisure or religious activities.

ii. Contributors to participatory research are neither full research collaborators, nor research participants. As such, they are not explicitly included in the protections offered to each of these populations. For example, because contributors to participatory research often take a role that more closely resembles field technicians than human research subjects, they are rarely given opportunity for informed consent, which requires clearly disclosing the risks and benefits of participation in a research project.

iii. Participatory research projects rely primarily on data contributed by private individuals to achieve their goals, and thus are understandably hesitant to risk alienating volunteers through under-considered or poorly documented practices. As Deutsch explains, "people's willingness to work on group problems and serve as part of a human sensor network will long-term depend on their ability to trust how their engagement and inputs to the project are handled" (Deutsch, 2013, p. 847).

iv. Some participatory research projects are conceived as "data collection by the people, for the people" (Robson, Kandel, Heer, & Pierce, 2011), designed with highly democratic aims. Project goals may include supporting a more informed and active citizenry, or social justice outcomes such as eliminating racial and ethnic inequality (Paulos et al., 2008; Cacari-Stone, Wallerstein, Garcia & Minkler, 2014), or addressing environmental concerns (Ottinger, 2009). To achieve these goals, transparency is often considered both a guiding principle and a necessary practice.

In participatory research, the needs to support privacy and to accurately document privacy-supporting practices are both practical and ethical concerns, creating tensions between privacy and transparency. But as our findings demonstrate, the need for complete and accurate policies is not always met. The goals of this study were twofold: understanding the current state of privacy policies in participatory research, and applying existing frameworks to develop concrete and actionable guidelines for ethically engaging volunteers in participatory research.

This work contributes an exploration of the data policies of two closely related participatory research paradigms—citizen science and participatory sensing—by analyzing the policies and, where possible, their related practices. This evaluation sheds light on the interdependency between technologies and data policies, relevant to many projects within human computation. Our evaluation also highlights discrepancies observed between policy and practice. To understand the ethical implications of our research, we draw on two sets of guidelines: The Belmont Report, advanced by the U.S. Department of Health and Human Services (HHS), and ethical guidelines for enabling privacy in participatory sensing (Shilton, Burke, Estrin, Govindan, & Kang, 2009). We conclude by synthesizing our research with these guidelines, providing a checklist of key privacy considerations for participatory research that may have value for other forms of human computation as well.

## 2. BACKGROUND

Here we define the key terms of privacy, personal data, and data policies, and then discuss the related phenomena of citizen science and participatory sensing, exploring why privacy is important to each.

Nissenbaum's theory of contextual integrity suggests that privacy concerns are formed and articulated in response to the values and norms of particular social context (Nissenbaum, 2011). Following researchers such as Shilton (2009), we define *privacy* in participatory research as: the right to manage access to voluntarily contributed personal data. Note that this excludes data collected involuntarily, such as that gathered through cookies and server logs, which is not a consideration unique to this context. We define *personal data* as data that contains identified or identifiable information (also referred to as PII) (Schwartz & Solove, 2014).

*Data policies* are collections of documents that dictate how one party, such as a project coordinator, may interact with another, such as a volunteer. While all data policies are intertwined, different types of policies may be grouped together and offered under a web page with a specific name. *Privacy policies* explain the protections projects take to support the privacy of volunteers and other users; *terms of use* describe how data is owned and licensed by projects and/or volunteers; *legal policies* clarify how projects and volunteers must comply with certain laws (Bowser, Wiggins, & Stevenson, 2013). Policies are typically enforced through a *user agreement*, or a legal contract that details the acceptable actions of projects and volunteers.

## 2.1. Citizen Science and Participatory Sensing

*Citizen science (CS)* is the involvement of volunteers in scientific research (Dickinson, Zuckerberg & Bonter, 2012). This strategy enables scientists to gather and analyze larger and more diverse data than would otherwise be feasible (Nov, Arzay, & Anderson, 2011). Volunteers derive pleasure from the topic being studied, social interaction with project leaders and volunteers, learning, and engagement with a community, process, or interface (Bowser et al., 2014). Citizen science projects are notable for diversity in activity structures and goals (Wiggins & Crowston, 2015). The data they produce have many uses; for example, eBird's avian species distribution data have been used to answer scientific research questions across several disciplines, resulting in over 90 peer-reviewed papers, and also to inform conservation policy (Sullivan et al., 2014).

By contrast, researchers define *participatory sensing (PS)* as a practice “which tasks everyday mobile devices, such as cellular phones, to form interactive, participatory sensor networks that enable public and professional users to gather, analyze, and share local knowledge” (Burke et al., 2006, p. 1). These applications are explicitly designed to provide direct utility to participants (e.g., by providing information streams such as data visualizations and reports) and also to external parties, such as researchers. Christin, Reinhardt, Kanhere, and Hollick classified PS into people-centric applications, which collect data about a person, and environmental sensing applications (2011). For example, PEIR is a people-centric application that uses GPS data from mobile phones to measure an individual's impact on and exposure to environmental contaminants (Mun et al., 2009). Noise-tube, on the other hand, is an environmental sensing application that collects data about environmental noise pollution (Maisonneuve, Stevens & Ochab, 2010).

We note that many researchers consider participatory sensing a subset of citizen science, or vice versa, or use these two terms interchangeably. However, like others (e.g., Heggen, 2013) we see a key difference in the role that technology plays in data collection; in contrast to most citizen science, which may be *supported* by new technologies, participatory sensing is *enabled* by interactive technologies such as mobile phones. As such, the field of participatory sensing is considerably newer, and at times as focused on the development of new technologies as on the research questions these technologies are designed to address. In addition, a majority of the projects included in our research had self-identified as either participatory sensing *or* citizen science. For these reasons, while noting that these definitions do overlap and also that additional categorizations of participatory research exist (e.g., “participatory urbanism,” Paulos, Honicky, & Hooker, 2009), we discuss citizen science and participatory sensing separately and identify recommendations suitable for both forms of participatory research.

## 2.2. Privacy Concerns in Participatory Research Projects

Privacy concerns in CS and PS are linked to both data collection and data sharing. In most CS projects, data is collected by discrete individuals and combined to create a larger data set. For example, Nature's Notebook[4] and Project Budburst[5] gather phenology data describing natural plant and animal life cycle events. These data become valuable in aggregate, and increase in value as data accumulate over time. These projects are typically designed to provide data for both scientists and other parties, such as policymakers. A second example, eBird, also provides volunteers access to raw data and visualization tools to reward and motivate participation (Wood et al., 2009).

In contrast, discrete individuals provide PS projects with data that is often automatically and/or continuously collected and streamed; for example, most seismographic sensors must record data without pause in order to identify both baseline activity and earthquake events. This data may be intended for a wide range of purposes, and may be shared between contributors, with society at large, with health care professionals, and with scientific researchers (Christin et al., 2011).

For both citizen science and participatory sensing, key privacy considerations arise around:

i. collection of sensitive location-based data (Shilton, 2009);
ii. collection of other sensitive data, such as health data or data about a protected species (Bohannon, 2009; Bowser et al., 2013); and
iii. accidental collection of data from secondary participants, such as those depicted in a photo (Henne, Szongott, & Smith, 2013).

These concerns are exacerbated by:

i. continuous data collection, which may enable the identification of not only discrete locations, but also routines (Christin et al., 2011; De Cristofaro & Henne, 2011);
ii. potential to combine data sets for re-identification of contributors (Bohannon, 2009); and
iii. norms of openness and data sharing prevalent in both communities, especially when raw data are shared (Kim, Mankoff & Paulos, 2013).

Most solutions to privacy concerns can generally be characterized as either *technical* or *policy-oriented*. Technical solutions are embedded in project implementation. Some are automatic safeguards built into data collection or retrieval, such as data perturbation or data cloaking (Christin et al., 2011). Others are activated as volunteers manually change their sharing permissions (Mun et al., 2009). Technical solutions do offer key protections, but may not offer complete protections, especially if volunteers do not understand how these solutions are implemented, and why.

Policy-oriented solutions support privacy primarily through project design, which can be extended to include user education via clearly communicating policies and the related risks and benefits. One simple example of a privacy-oriented solution is anonymization, which can occur before

---

[4] https://www.usanpn.org/natures_notebook

[5] http://www.budburst.org

(e.g., through the use of an alias) or after data collection (e.g., by purging PII), but may be at odds with research goals. Some policy-oriented solutions may be more flexible than technical solutions. For example, volunteers may simply be instructed to register for a project anonymously, by selecting a user name that does not include their full name, email address, or other PII.

There are some existing resources for supporting privacy in the creation of new technologies. Deutsch (2013) suggests the Privacy by Design[6] guidelines for human computation researchers. While these guidelines are beneficial, we note that with one exception ("visibility and transparency") they are largely oriented towards technical and not policy-oriented solutions. A second set of guidelines, which encompasses both technical and policy-oriented solutions, is the Fair Information Practices[7] published by the Federal Trade Commission (FTC). These generalized guidelines are designed to inform design in a broad range of contexts. Building off this starting point, Shilton, Burke, Estrin, Govindan, & Kang (2009) identified three design principles to ensure privacy in participatory sensing:

i. *Participant primacy* asserts that data collection should always be done with an individual's knowledge and explicit agreement.

ii. *Data legibility* suggests that volunteers should understand how their data are used by researchers, and be supported in their own data analysis.

iii. *Longitudinal engagement* represents volunteers' ability to change their data sharing permissions over time as contexts change.

Our analysis integrates these guidelines with the basic ethical principles articulated in The Belmont Report (U.S. Department of Health and Human Services, 1979), and the core of the ethics framework that informs human subjects research review and regulation in the U.S.:

i. *Respect for persons* asserts that humans are autonomous agents, with the right to decide what happens to them.

ii. *Beneficence* means that participants' well-being should be actively supported and participants should be protected from harm.

iii. *Justice* requires equitable distribution of research risks and benefits so that those who are exposed to greater risks also reap greater benefits.

We present our evaluation of the intersections between both sets of guidelines in Table 1. Drawing on the combined ethical guidelines allows us to offer a practical checklist of considerations to inform the design of participatory research projects, derived from a framework presented below. We expand on this table in our discussion, contextualizing the intersections of these two sets of

---

[6] http://www.privacybydesign.ca/

[7] http://www.ftc.gov/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission

guidelines within our research results to construct a checklist of ethical principles for participatory research design.

| **Belmont Report Principle** | **Design principle from Shilton et al. (2009)** | **Intersections** |
|---|---|---|
| Respect for persons | All, but especially participant primacy | *Respect for persons* and *participant primacy* suggest that participants have the right to determine what happens to them through informed consent. *Data legibility* and *longitudinal engagement* advocate that volunteers should be continually informed about the uses of their data and their rights as contributors. |
| Beneficence | Data legibility | *Beneficence* suggests that benefits to participation should be maximized, and *data legibility* suggests that participants may benefit from access to data and analysis opportunities.. |
| Justice | Longitudinal engagement | *Justice* advocates that risks and benefits should be balanced. *Longitudinal engagement* suggests that volunteers should continually evaluate benefits and risks, which change based on context. |

Table 1. Intersections between the design principles from Shilton et al. (2009) and the Belmont Report.

## 3. STUDY DESIGN

We used a document analysis process to analyze participatory research projects' policies (Bowen, 2009), which we collected from the Internet. We inductively generated a coding schema for analyzing the policies of citizen science and participatory sensing projects. We then used theoretical sampling to select a diverse range of projects, retrieved their online policies, and applied the classification schema to identify privacy issues. Note that our aim was to understand, generally, the range of policies—particularly those related to privacy—present and absent in participatory research projects. We therefore analyzed policies from a broad range of projects with diverse organizational arrangements that were expected to prompt variations in policies (e.g., due to sector-specific regulatory requirements).

### 3.1. Schema Development

Since there were no suitable classification schemes in the related literature, we developed a schema for evaluating data use policies, ownership claims, and privacy considerations related to data contributed by volunteers. The schema was generated via inductive content analysis of policies drawn from twenty citizen science projects. We first selected a landscape sample (Bos et al., 2007) to include projects of diverse sizes, types of data collected, and methods of data collection. This form of theoretical sampling maximizes diversity to enable a broad view of the phenomenon, and does not seek statistical representativeness.

We visited the websites of each project, retrieving all text that could potentially be considered "policy". Using a Thematic Analysis approach (Braun & Clarke, 2006), we searched across the data to identify snippets of policy with common themes. These themes were organized into a classification scheme, with high-level categories that included: presence and location of policy statements within websites; elements of legal and privacy policies, terms of use, and user agreements; types of data collected upon registration and during participation; and intellectual property claims. Independent experts in information policy and intellectual property law reviewed the schema to verify its completeness and congruence with North American legal definitions.

## 3.2. Sampling and Data Collection

We selected 15 citizen science projects and 15 participatory sensing projects for comparison of their policies, again using a landscape sampling technique. Due to scoping limitations and linguistic concerns, all of our projects had a strong presence in North America, impacting the legal paradigms and specific considerations applicable to each project. This is reflected in our reporting and discussion of the research results, where we focus primarily on U.S. laws and policies. Broadening this analysis to adequately integrate the more comprehensive and complex privacy protections of other countries would be a valuable direction for future research.

CS projects were selected for variety in scientific focus, organizational structure, scale of operations, and range of supporting technologies, using criteria from Wiggins (2013). These projects were not part of the sample used to generate the classification scheme. We sampled CS projects that collected observational data, as these data are typically geo-located and have potential to engender strong privacy concerns. Volunteers in the projects we sampled collected data related to topics including plants, animals, invasive species, and weather. Our sample of CS projects included projects following contributory, collaborative, and co-created models (Shirk et al., 2012), with goals including investigation, conservation, and education (Wiggins & Crowston, 2011).

Sampling from the wider population of active projects is feasible in CS (Kim, Mankoff, & Paulos, 2013; Wiggins & Crowston, 2014), but of 31 PS projects reviewed in 2011 by Christin et al., only four could be conclusively identified as active in August 2013. PS projects were therefore selected opportunistically, with emphasis on active projects, as there are far fewer readily identifiable PS endeavors (due to their scarcity, PS projects were excluded from the sample used for schema development). PS projects collected data related to air quality, water quality, seismology, mobile connectivity, and volunteered geographic information. These projects followed contributory, collaborative, and collegial models of participation (Shirk et al., 2012) with primary goals characterized as action, investigation, conservation, virtual, and education oriented (Wiggins & Crowston, 2011).

Both CS and PS projects may face significant resource constraints that can impact their performance as relates to this study. While we do not disclose the names of specific projects we surveyed in recognition of these limitations, additional information about the sample may be found in the appendix.

In addition, we note that volunteers were not the subjects of research in these projects, but acted as collaborators in a role similar to field technicians, as previously discussed. Based on close scrutiny of all public-facing materials for each of the sampled projects, we verified that none were designed to conduct human subjects research as defined by HHS. For this reason, we did not further pursue questions around informed consent, nor did we attempt to determine whether projects went through an IRB approval process.

## 3.3. Analysis Procedure

Two researchers independently retrieved policies from the websites of each project to form a comprehensive corpus for document analysis. The corpus included content that explicitly described policies, such as content labeled as “Privacy Policy” or “Terms of Use.” It also included snippets of policy found on pages that declared a different primary purpose, e.g., “About” or “FAQs.”

Using a subset of the full classification scheme for analysis of privacy-related policy elements (as opposed to, for example, policies on intellectual property or liability), both researchers independently coded the corpus. Upon review, we clarified definitions and argued to consensus; there were only two instances (of 550) in which a final determination could not be made because the language of the policy could be interpreted in multiple ways.

We then returned to each project to record objective indicators of privacy practices, verifying which items of personally identifiable information were requested upon registration. Collecting these data proved challenging, as some projects lacked a formal online registration processes. Due to practical challenges in achieving a suitable level of completeness for comparison to stated policies, we report these practices in the form of narrative evidence.

# 4. RESULTS

In keeping with the goal of understanding the current range of privacy policies as stated and enacted, our qualitative analysis yielded primarily descriptive results. We first describe general demographic characteristics of the sampled projects, followed by details of how the information they collected was commonly described and used. In general, we observed substantial variety along each dimension, with policies ranging from well-crafted and thorough to vague to entirely absent. An overview of key findings is presented in Table 2. We also examine role-specific data management permissions, and briefly discuss compliance with key U.S. privacy laws such as COPPA and CAN-SPAM. Similar diversity was observed in these categories of policies as well.

## 4.1. Sample Description

Fifteen (of 15) citizen science projects and 10 (of 15) participatory sensing projects were classified as stable and active. Five PS projects were classified as prototypes; of these, two were considered abandoned. Of the 30 projects surveyed, only one project had no online policies whatsoever. The majority of projects posted policies labeled as “terms of use” and “privacy policy.” Legal policies, including references to national, state, or local laws, were less common. Most projects hosted elements of a user agreement, though few provided a complete user agreement, as discussed below.

The web pages where we identified elements of policy were not always obvious places to seek policy information. We considered this important, because users may reasonably expect all policies to be labeled as such, and might not expect important policies to be mentioned exclusively on “About” or “FAQ” pages. Several projects hosted policies in multiple locations, including pages for submitting data, describing group membership, listing research results or publications, hosting training resources, describing technologies, and describing the project team’s research. In some cases, key policies were presented at the points of the participation process in which they were most relevant to users.

| **Policy aspect** | **Summary of findings** |
|---|---|
| Information collected during registration | Projects collected information including full name, email address, username, demographic data, organizational affiliation, self-reported level of expertise, mobile device model, and health data. Information was shared with the project team, project affiliates, other project participants, and advertisers. |
| Information collected during participation | Projects collected information including latitude and longitude, observation date or time, descriptive data, and images. Information was shared with the project team, scientists or researchers, other project participants, and advertisers.<br>Some projects collected potentially personal information, such as location, without explicitly mentioning it in their policies. |
| Data management permissions | Project managers may edit, modify, or delete volunteer data, or hide data from the public view.<br>While few projects explicitly describe volunteer permissions, many projects enabled volunteers to selectively hide data from public view. |
| Legal policies | A number of projects mentioned compliance with COPPA or CAN-SPAM. Federal projects also mentiond FOIA, and some projects described compliance with EU privacy laws. |

Table 2. Summary of findings for four key aspects of data policy.

Some projects offered contact information for volunteers with questions about policies. For example, most provided a contact form, and some offered either a generic email account or a specific individual's email or phone number. Providing contact information in this manner may either imply recognition that policies may be difficult to comprehend or compliance with institutional mandates (e.g., for federal agencies). A few projects did not provide contact information, while others offered multiple modes of contact within their policy statements, demonstrating a wide range of practices.

## 4.2. Information Collected During Registration

Ensuring that participants understand the types of information collected and shared is a pre-requisite for supporting informed participation. As expected, a number of policies listed the types of information collected during registration. Note that because these policies described information about participants collected *prior* to any data submission, a volunteer who registered for a project but never contributed data could still have personal information shared in the ways discussed below. The types of information collected included: full name, email address, mailing address, username, and demographic data. Two CS projects also mentioned collecting information on organizational affiliation or self-reported level of expertise, while two PS projects reported collecting mobile device model in order to standardize sensor data. One personal health monitoring site asked for weight and height, which were used to personalize data services for participants. A few projects of each type failed to mention any types of information collected during registration.

Some projects reserved the right to share the information collected during registration with parties that included the project team, project affiliates, other project participants, and advertisers. A few CS projects (but no PS projects) explicitly stated that this information would never be shared, while others offered more generalized assurances. A few CS projects and some PS projects did not describe their information sharing practices.

According to policies, participant information could be shared for many reasons, including website operations, announcements, advertising, and data verification. Again, some projects also provided generalized statements, such as promising that email contact will be "kept to a minimum." A few CS projects and some PS projects also did not mention the conditions under which information collected during registration may be shared. An often overlooked and seemingly innocuous category of privacy-related policy, these data sharing and usage policies were the same type as those criticized in the Facebook emotional contagion study mentioned in the introduction, which many users and critics viewed as sanctioning ethically questionable practices that violated user expectations.

## 4.3. Information Collected During Participation

Some projects also described the types of data collected through observation or sensing, including: latitude and longitude, observation date or time, descriptive data, and images. Some CS projects and a few PS projects made no mention of types of information collected, but these details may have been listed elsewhere, e.g., in instructions. Data collected through observation or sensing was reported as shared with parties including the project team, scientists or researchers, other project participants, and/or advertisers. As with other types of information, some projects did not mention what data were shared through observation or sensing.

In contrast to policies describing the *sharing* of contact information, policies describing *access* to contributed data were slightly less common. This may be because some sharing was implicitly assumed; for example, the value proposition for citizen science is generally predicated on sharing volunteer-generated data with researchers and decision-makers. Even if that sharing were limited to the project team, it could still imply a sizable network of staff and researchers (e.g., Sullivan et al., 2014).

Since detailed study of actual practices was not feasible in the cases of prototypes or abandoned projects, our comparison of stated policy with practice was thorough but not exhaustive. Still, a few obvious discrepancies emerged. Significantly fewer projects mentioned collecting location, observation, and descriptive data than actually gathered these types of data. As with registration, participants would discover which data were collected once they started contributing, but the key ethical consideration here is providing information for informed decision-making about participation prior to engaging. Due to our sampling and the nature of the phenomena, virtually all of the projects we surveyed collected location-based data, but this was highlighted in policies by only a few[8]. Likewise, most projects collected date and time information, and most CS projects also collected descriptive data about volunteer observations, but this was only mentioned occasionally.

All of these seemingly trivial details have the potential to be personally revealing under the right circumstances. When the risks of participation are unmentioned, volunteers may proceed on unfounded assumptions and later find themselves regretting their decision to contribute.

## 4.4. Data Management Permissions

Following our definition of privacy as “the right to manage access to voluntarily submitted data,” understanding role-related permissions for data management (such as the rights to add, modify, or remove data) is key to supporting volunteer privacy in the context of participatory research projects.

Some policies described the permissions granted to projects, asserting the rights of project managers to edit or modify volunteer data, delete volunteer data, and hide data from the public view. Some policies also explained the permissions granted to volunteers. In most cases, policies simply did not mention permissions at all. This might reflect an implicit assumption that projects own the data submitted by volunteers, which conflicts with the default intellectual property rights granted under U.S. copyright law (Scassa & Chung, 2015).

Volunteers’ permissions for data management were confirmed for most projects by reviewing data submission forms. A larger number of projects in both CS and PS allowed their volunteers to selectively hide data from public view than advertised the option in their policies. For example, one project allowed users to convert a precise location to a random location within a certain radius, but did not describe this feature prior to data submission. This type of functionality supports both participant choice and privacy protections, demonstrating a technological implementation of policies that emphasize respect for persons and participant primacy. Ensuring that participants are

---

[8] At least one project has since developed a proactive privacy alert for new users that explains the risks of openly sharing geolocated data and describes the features that users can employ to protect their privacy.

aware of this functionality prior to contributing data further demonstrates this respect and consideration.

## 4.5. Legal Policies

While an interpretation of legal policy is beyond the scope of this paper, the data policies we analyzed indicated an awareness of two U.S. laws relating to privacy: the Children's Online Privacy Protection Act (COPPA), and the Controlling the Assault of Non-Solicited Pornography and Marking (CAN-SPAM) Act.

COPPA protects the online privacy of children under the age of 13. All for-profit websites collecting personally identifiable information from children are legally required to comply with COPPA[9]. Many other websites, such as those sponsored by federal agencies, comply as well, either due to agency regulation or as a best practice.

Some projects included language in their data policies about COPPA compliance, but with widely variable interpretations. Some linked to COPPA from their policies; others mentioned that children under 13 must obtain permission from their parents; and several specifically noted that their project did not knowingly or intentionally collect data from minors. Still others simply stated, "[project] complies with COPPA." Unfortunately, none of these projects demonstrated full COPPA compliance per the FTC. The nearest approximation came from a project operated by a federal agency; as such, its policies were developed with legal consultation and subject to extensive review processes, but still missed two key points. These findings indicate that most projects were at least aware of COPPA and intended to comply, even when they were technically exempt from it, but nonetheless fell short.

The Controlling the Assault of Non-Solicited Pornography And Marketing Act (CAN-SPAM) established United States standards for sending commercial emails. While CAN-SPAM compliance is complex, a major and well-known component is that email recipients must be allowed to opt out of receiving advertisements. As mentioned earlier, a few CS and PS projects shared content for purposes of advertising; some of these included instructions for opting out of receiving emails in their data policies. Interestingly, more projects described how to opt out of receiving promotional emails than informed volunteers that email addresses were shared for such purposes in the first place.

Three additional legal policies were mentioned by a few projects. Some federally-sponsored CS projects noted that data submitted by volunteers becomes part of agency record, and is therefore subject to Freedom of Information Act (FOIA) requests, meaning that data records containing potentially sensitive information could be made publicly available A few projects discussed compliance with the European ePrivacy Directive, which outlines conditions for collecting information through cookies and web server logs. Finally, a few PS projects mentioned US-EU Safe Harbor, which defines how U.S.-based companies can comply with much stricter European Union regulations on personal data collection.

---

[9] http://www.coppa.org/comply.htm

# 5. DISCUSSION

This section reviews the data policies we found and then discusses the ethical implications of our findings. We also propose a checklist of ethical practices for participatory research design, drawing from the intersections of the Belmont Report and the guidelines advanced by Shilton et al. (2009).

## 5.1. Summary of Key Findings

Our research indicates that practices to implement data policies lag behind the ideal. While most projects surveyed had *some* data policies, only some CS and fewer PS projects included elements of every type of policy from our classification scheme: terms of use, privacy policies, legal policies, and user agreements. This is consistent with the findings of Bohémier and colleagues (2010), who searched for six types of data policies and reported an average of 3.8 types of policies implemented by organizations in their sample. These researchers also mentioned difficulty retrieving data policies, noting that not all policies were publicly available. Our findings reinforce these concerns and suggest that conditions have not improved in participatory research projects despite the dramatic increase in public attention to personal privacy concerns in the intervening years.

A majority of the projects included in our sample had published web content that we considered privacy policy, and a slightly different majority had published content that met our definition of terms of use. A smaller subset of projects actually hosted their privacy policies on a clearly labeled “privacy policy” page, and less than half of projects hosted their terms of use on a clearly labeled page. Similarly, while about half of all projects included elements of legal policies or user agreements, just two included a clearly labeled “legal policies” page and only one hosted a page labeled “user agreement.” Other fragments of policies were distributed across numerous pages.

## 5.2. Ethical Implications

Several ethical implications emerge from reviewing these findings through the lens of two frameworks: the Belmont Report and the design principles advanced by Shilton et al. (2009). The Belmont Report’s principle of *respect for persons* suggests that research participants must be truly informed regarding the risks and benefits involved in participation. Similarly, *participant primacy* advocates that data collection should be conducted with participants’ knowledge and explicit agreement (Shilton et al., 2009). We found that most projects did not adequately inform participants of the types of information collected during registration and data collection nor how this information may be used. They either omitted such information or presented policies in ways that made them difficult to find. These practices are contrary to respect for persons and participant primacy.

Transparency can also be a legal concern. For example, a number of U.S. laws dictate how participatory data may be collected by federal agencies, which operate a growing portfolio of CS projects and are increasingly involved in PS as well. The Paperwork Reduction Act (PRA) requires each agency to undergo a lengthy review process by the Office of Management and Budget, during which project leaders must describe exactly which types of data will be collected and how. These details must also be publicly reported in the Federal Register[10] prior to beginning data collection. The Freedom of Information Act (FOIA), designed to increase citizens' trust in government through transparency, provides a mechanism for public access to agency records. Laws such as these, designed to protect the public good, complicate the implementation of participatory research in federal agencies by requiring a substantial up-front investment in compliance and reducing flexibility to adapt projects to better meet participant and research needs.

The Belmont Report principle of *justice* suggests that the benefits of research should outweigh the risks. When this principle is combined with *longitudinal engagement* (Shilton et al., 2009), it becomes clear that the negotiation of risks and benefits is not a one-time decision, but a complex equation with constantly shifting variables. Instead of conveniently ignoring these very real risks, projects could use their policies to educate volunteers about the potential privacy risks associated with location-based data, how these risks may change between contexts and over time, and how to best address such concerns within the specific context of the project. To be most effective, the information provided through policies can be coupled with technological safeguards, for example, through design that supports the sharing of location data only at the user's discretion, (e.g., De Cristofaro & Soriente, 2011).

The Belmont Report's principle of *beneficence* asserts that benefits should be actively maximized. The design guideline of *data legibility* similarly suggests that participants may benefit from participating in analysis and interpretation in addition to data collection (Shilton et al., 2009). Some projects already provide tools to support more extended forms of participation. Volunteers working with eBird, a citizen science project collecting avian species occurrence data, are often avid birders who care about both bird conservation and social recognition. The project leaders have developed a suite of features to support birding community practices, including basic tools for exploring their own data and data collected by others, which involves a level of social transparency as to the activities of other participants (Sullivan et al., 2014). While these tools are strongly desired by users, they simultaneously pose a potential threat to participant privacy, creating an unavoidable tension that requires ongoing monitoring and management. For volunteers to reap the greatest benefits from participation, such features should be clearly described by the projects that support them. Synthesized from our empirical findings and based on our contextualization of the combined ethical principles, Table 3 identifies seven Ethical Principles for Participatory Research Design.

- *Ethical Engagement*: projects should identify core ethical principles—based on established guidelines and/or unique project goals—to govern design and implementation. These principles should be considered during the development process for every human computation system, due to the fundamental reliance on human participation, and posted alongside project policies.

---

[10] https://www.federalregister.gov/

| **Belmont Report Principle** | **Design principle from Shilton et al. (2009)** | **Intersections** | **Ethical Practices for Participatory Research Design** |
|---|---|---|---|
| All | All | *The Belmont Report* advances generalized ethical practices for engaging humans in research; Shilton et. al's (2009) design principles offer concrete suggestions for supporting privacy in participatory research. | *Ethical engagement:* Projects should identify a set of ethical principles—based on established guidelines and/or unique project goals—to govern project design and implementation decisions.<br>*Ongoing assessment:* Stakeholders including funders, project leads, developers, and volunteers should continually evaluate adherence to these principles. |
| Respect for persons | All, but especially participant primacy | *Respect for persons* and *participant primacy* suggest that participants have the right to autonomy through explicit and informed consent. *Data legibility* and *longitudinal engagement* advocate continually informing volunteers about data uses and their rights as contributors. | *Informed participation:* Projects should provide volunteers a clear and complete explanation of participation requirements prior to accepting contributions, e.g. via a process of informed consent.<br>*Evolving consent*: If the terms of participation change, projects should inform volunteers and repeat the process of informed consent. |
| Beneficence | Data legibility | *Beneficence* suggests that benefits to participants should be maximized, and *data legibility* suggests that participants may benefit from access to data and analysis opportunities. | *Participant benefit:* Projects should maximize the benefits to individuals and/or groups based on the desires of their target volunteers. |
| Justice | Longitudinal engagement | *Justice* advocates that risks and benefits should be balanced.<br>*Longitudinal engagement* suggests that volunteers should be able to continually evaluate benefits and risks, that change based on context. | *Meaningful choice:* Projects should present risks and benefits associated with participation to enable meaningful participation choices.<br>*Evolving choice*: Projects should enable volunteers to make meaningful choices about participation as contexts change. |

Table 3. Ethical Principles for Participatory Research Design, derived from the intersections of the Belmont Report and guidelines for protecting privacy in participatory sensing (Shilton et al., 2009).

- *Ongoing Assessment:* stakeholders—including funders, project leads, developers, and volunteers—should regularly evaluate adherence to these principles. Holding both projects and volunteers accountable for supporting ethical participation may encourage mutually beneficial co-ownership of a project and its goals.
- *Informed Participation*: projects should provide a clear and complete explanation of participation prior to accepting contributions tied to user identities or locations, via an explicit informed consent process. Informed consent must be clear and thorough but may vary in level of formality depending on context. To effectively support agency, all policies should be hosted on clearly labeled pages and explained at in-person recruitment and training events. Projects should also provide information on relevant legal compliance.
- *Evolving Consent*: if the terms of participation change, policies should be promptly updated and projects should both inform volunteers and repeat the process of informed consent. To facilitate understanding of the changes and their implications for participants, the former policies can be paired with new guidelines and examples of how the new policies apply in policy update communications sent to ongoing volunteers.
- *Participant Benefit:* projects should maximize benefits to individual participants and/or social groups based on their unique interests. Adopting a user-centered design process can ensure that such benefits are built into the system from the earliest stages of design and implementation, in addition to providing opportunities to identify and rectify problematic language in policy statements. Describing participant benefits in policies can also serve as a recruitment tool, as discussed below.
- *Meaningful Choice*: projects should present the risks and benefits associated with participation to enable volunteers to make meaningful choices on how exactly to participate based on their personal preferences. This principle extends *Informed Participation* by allowing volunteers who have already agreed to contribute to structure their activities based on their own assessments of key benefits and risks. For example, a volunteer may choose to disclose personal information in return for a perceived benefit, such as the potential for social interaction with other community members.
- *Evolving Choice*: participant choice should be supported not just once, but multiple times, particularly in response to changing contextual conditions. For example, a volunteer may choose to disclose personal information in return for a perceived benefit in one geographic location (e.g. a public space like a national park), but decline to disclose the same information in another (e.g. a private home). Clear data policies will support both meaningful choice and evolving choice.

These seven principles are offered as guidelines to inform the development of policies and the design of technologies supporting participatory research. Upholding ethical best practices will support and enhance the reputations of both individual projects and the field of human computation more broadly. In addition, these practices can support stronger relationships between projects and their volunteers, and between different projects.

## 5.3. Implications for Relationships Between Projects and Volunteers

For diverse forms of human computation, participant motivation—which influences the recruitment and retention of contributors—is a key concern. The quality and frequency of communication between project leaders and contributors is one factor supporting volunteer recruitment and retention (Bowser et al., 2014; Rotman, et al., 2012). Transparency about the information being collected and its uses therefore achieves a strong ethical position (*informed participation; meaningful choice*) while also supporting practical needs via development of respectful relationships with volunteers that can further sustain engagement. Similarly, projects that offer multiple options for information sharing, e.g., allowing volunteers to selectively hide their location from public view, are acting both ethically and pragmatically by ensuring that volunteers feel comfortable with the visibility of their contributions (*evolving choice*). Finally, providing resources that help meet volunteers' goals, such as the birder-centric tools offered by eBird, demonstrates care and respect extended by project leaders to volunteers (*participant benefit*) that can otherwise be difficult to communicate.

In some cases, common project evaluation practices, such as gathering contributor feedback, can be considered human subjects research. It appears, however, that project leaders frequently overlook the potential need for human subjects ethics review and associated protective measures in participatory research; this is likely because humans are not the primary focus of their data collection activities. But in addition to explicitly considering whether human subjects review is appropriate, ensuring *informed participation* may be a practical mechanism for proactively managing privacy concerns while promoting volunteer trust. The technological implementation of an informed consent policy could employ an actual informed consent form, or could present a clear, short user agreement with an obvious opt-in checkbox before submitting data. As these examples demonstrate, many ethical practices can also be considered best practices for project design that should enhance overall project performance.

## 5.4. Implications for Relationships Between Projects and Partners

Data policies can facilitate data sharing and integration between and among participatory research projects and related parties by establishing the rules of co-engagement. If the policies of two or more projects are compared before entering into a partnership, potential conflicts can be proactively identified before they materialize. Up-front negotiation of strategies that fulfill the intentions of each set of policies can then commence with a full understanding of key details to resolve. For example, a pair of citizen science projects with different policies regarding the visibility of volunteer PII or location information might choose to follow the more conservative set of policies in their joint effort for consistency with the stated ethical principles of the more conservative project (*ethical engagement)*. They might alternately enable an opt-in selection with clearly worded explanations of the risks and benefits so that volunteers can make informed decisions about contributing data to joint efforts (*evolving consent)*. Collaborating projects could also agree to hold each other accountable for adhering to their stated ethical principles (*ongoing assessment).*

Every project leadership team is constrained in the extent to which they can support other projects' needs and ethical principles in data sharing arrangements. Providing direct access to live databases, for example, requires much stronger trust (and possibly legal agreements such as a Memorandum of Understanding) between partners than providing a curated data set in which volunteers' identifying information has been removed or obscured. The effort required to harmonize decisions across projects, especially related to volunteers' privacy, sometimes exceeds this limit and serves as a reminder that sharing and repurposing data often requires that the data be curated specifically for those uses.

## 6. FUTURE WORK AND CONCLUSION

Through description, analysis, and discussion, this study advances our understanding of the current state of data policies, especially privacy policies, and related practices in participatory research projects, a growing area of interest within human computation. We examined our findings within the context of two sets of ethical guidelines, which we integrated to contribute an initial framework of Ethical Principles for Participatory Research Design.

This study has a number of limitations. The methods and data cannot explain the sources of the shortcomings we identified, but point to a few likely reasons for the current condition of data policies in participatory research. Projects without data policies may not fully understand what data policies are or why they are needed. Researchers from the Center for Embedded Network Sensing at UCLA noted that "the culture of ad hoc planning is very strong in some of the field-based sciences" (Wallis, Mayernik, Borgman, & Pepe, 2010) and our observations could be construed as further evidence of this claim. While ad hoc planning is sometimes necessary to address unexpected challenges, a lack of formal planning can also suggest limited resources (Wiggins, 2013). However, the presence of any data policies indicates some level of consensus that such policies are important.

For all of these reasons, we believe that the fragmented policies we observed result from the on-the-ground realities of project design and operation that are often antithetical to idealized top-down approaches to policy development. In other words, project leaders likely understand the importance of these policies but lack the resources to develop or uphold them, and consider it a less urgent issue than other operational needs. Anecdotally, we also saw the most comprehensive policies coming from the projects with high levels of participation or strong institutional support. This suggests a potential learning effect, e.g., these projects are more likely to have encountered incidents that prompted them to develop thorough policies. We believe that future work investigating the rationale behind project leaders' policy management strategies, and also examining the influence of changing policies, would be valuable for developing stronger tools and standards to ensure ethical practices.

Throughout this study, we observed that projects operated by federal agencies typically demonstrated much more effort to thoroughly comply with policies and laws intended to safeguard the public. The mechanisms through which such compliance is ensured, however, represent substantial time and effort that can slow initial project development. These projects provided an example of why the challenges of self-regulation may be preferable to becoming the target of new policies that could introduce similar roadblocks for human computation researchers and project leaders. As granting agencies become more particular about data management practices, open access to research results, and transparency of research funded by public monies, new requirements related to data policies could foreseeably emerge.

We also recognize that several of our suggestions may be difficult for some projects to implement. For example, not all projects require volunteers to register before contributing data, removing an obvious method for securing *informed participation*. Principles like *informed participation* and *evolving consent* could still be supported by designing workflows that require participants to click through a set of instructions or an annotated introductory activity in the process of initial engagement. Recognizing these limitations, our Ethical Principles for Participatory Research Design are not intended for strict compliance "to the letter", but (like any set of ethical guidelines) must be interpreted "in spirit". Prior research has shown that participatory research projects are often constrained by limited time, resources, technology, and personnel (Kim, Mankoff, & Paulos, 2013; Wiggins, 2013). We advocate for practical solutions, using both policies and technologies, to protect participant privacy to the fullest extent reasonable, and to clearly communicate the benefits and risks associated with participation. Communication is, of course, an issue of education as well as one of policy. While investigating strategies for educating participants was beyond the scope of this work, education remains a key consideration in participatory research project design. Anecdotally, we observed few efforts along these lines in the projects we examined.

Finally, we focused our investigation on participatory research, but note that there are significant shared ethical concerns in other models of human computation. Irani and Silberman, who studied worker invisibility on Amazon's Mechanical Turk, presented a new tool to support transparency in crowdworking, exemplifying how the principle of respect for persons (and participant primacy) might be applied (Irani & Silberman, 2013). Future research would clearly benefit from examining the data policies of models of human computation beyond participatory research.

Our research suggests that while the majority of these participatory research projects demonstrated a basic understanding of the need for data policies, many had incomplete policies or inaccurately described their actual practices. These issues have significant ethical implications, discussed through the dual lenses of the principles of *respect for persons*, *beneficence*, and *justice* articulated in the Belmont Report, as well as the design principles of *participant primacy*, *data legibility*, and *longitudinal engagement* (Shilton et al., 2009). By combining these sources, we advance seven Ethical Principles for Participatory Research Design: *ethical engagement, ongoing assessment, informed participation, evolving consent, participant benefit, meaningful choice*, and *evolving choice*. We argue that in many cases, ethical practices are also best practices, especially as these facilitate good relationships between projects and their volunteers, and between different projects seeking to share data or otherwise collaborate. We believe that by establishing strong, ethically-grounded project-level policies, the research community can be better prepared to inspire, rather than react to, potential future research policy changes to support the rights of participants in human computation projects more broadly.

In addition, our findings point to a need for more deliberate alignment between policies and technical solutions. We believe that it would be advantageous to design and revise policies and technologies in parallel, with deliberate efforts to ensure that the people writing the policies are in communication with those who are constructing the technologies (a perspective inspired by Shilton, 2012). The nature of the policy-technology misalignments for many projects in our sample also suggests that any time participation processes are modified or updated, it is an opportune time to review policies to ensure coherence with new procedures.

The responsibility for ethical engagement in participatory research lies almost entirely with those who design and build human computation systems. Therefore, we suggest that when establishing or revising project policies and practices, project leaders should incorporate the Ethical Principles for Participatory Research Design into their decision-making processes. Working with these principles will remind both project leaders and technology developers of the fundamental importance of respectful relationships with volunteers, without whom the project goals cannot be achieved.

## 7. ACKNOWLEDGMENTS

This work was supported by NSF Grant ACI 0830944, the Cornell Lab of Ornithology, and the Alfred P. Sloan Foundation. We thank Rob Stevenson, Bob Gellman, Everett Wiggins, and anonymous reviewers for their input on early drafts.

## 8. REFERENCES

Bohannon, J. (2013). Genealogy databases enable naming of anonymous DNA donors. *Science*, 339(262). DOI: 10.1126/science.339.6117.262.

Bohemiér, K., Atwood, T., Kuehn, A. & Qin, J. (2011). A content analysis of institutional data policies. In *Proceedings of the 2011 Joint International Conference on Digital Libraries,* Ottowa, Canada. DOI: 10.1145/1998076.1998159.

Bos, N., Zimmerman, A., Olson, J. Yew, J., Yerkie, J., & Dahl, E (2007). From shared databases to communities of practice: A taxonomy of collaboratories. *Journal of Computer-Mediated Communication, 12(*2), 652–672. DOI: 10.1111/j.2083-6101.2007.00343.x.

Bowen, G.A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27-40. DOI: 10.3316/QRJ0902027.

Bowser, A., He, Y., Rotman, D., Preece, J., Hansen, D., Boston, C., & Hammock, J. (2014). Motivating participation in citizen science. In *Proceedings of the European Conference on Social Media,* Brighton, UK.

Bowser, A., Wiggins, W. & Stevenson, R. (2013). *Data Policies for Public Participation in Scientific Research: A Primer*. DataONE: Albuquerque, NM.

Bowser, A., Wiggins, W., Shanley, L., Preece, J., & Henderson, S. (2013). Sharing data while protecting privacy in citizen science. *ACM Interactions, 21(*1), 70-73. DOI: 10.1145/2540032.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. DOI: 10.1191/1478088706qp063oa.

Burke, J., Estrin, D., Hansen, M. Parker, A., Ramanathan, N., Reddy, S., & Srivastava, M. B. (2006, October). Participatory sensing. In *Proceedings of the 4th ACM Conference on Embedded Sensor Systems*, Boulder, CO.

Cacari-Stone, L., Wallerstein, N., Garcia, A., & Minkler, M. (2014). The promise of community-based participatory research for health equity: A conceptual model for bridging evidence with policy. *American Journal of Public Health,* 104(9), 1615-1623. DOI: 10.2105/AJPH.2014.301961.

Christin, D., Reinhardt, A., Kanhere, S., & Hollick, M. (2011). A survey on privacy in mobile participatory sensing applications. *Journal of Systems & Software, 84*(11), 1928–1946. DOI: 10.1016/j.jss.2011.06.073.

De Cristofaro, E., & Soriente, C. (2011). PEPSI: Privacy-Enhanced participatory sensing infrastructure. In *Proceedings of the Fourth ACM Conference on Wireless Network Security*, Hamburg, Germany. DOI: 10.1109/TIFS. 2013.2287092.

Deutsch, J. (2013). Data, dopamine, and privacy by design. In P. Michelucci (Ed.), *Handbook of Human Computation* (847-855). New York, NY: Springer.

Dickinson, J., Zuckerberg, B., & Bonter, D. (2010). Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, & Systematics, 41*, 149-172. DOI: 10.1146/annurev-ecolsys-102209-144636.

Heggen, S. (2013). Participatory sensing: Repurposing a scientific tool for STEM education. *ACM Interactions, 20*(I), 18-21. DOI: 10.1145/2405716.2405722.

Henne, B., Szongott, C., & Smith, M. (2013). SnapMe if you can: Privacy threats of other peoples' geo-tagged media and what we can do about it. In *Proceedings of the 6th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, Budapest, Hungary. DOI: 10.1145/2462096.2462113.

Irani, L., & Silberman, M. (2013). Turkopticon: Interrupting worker invisibility on Mechanical Turk. In *Proceedings of the 2013 ACM Conference on Human Factors in Computing Systems*, Paris, France. DOI: 10.1145/2470654.2470742.

Kim, S., Mankoff, J., & Paulos, E. (2013). Sensr: evaluating a flexible framework for authoring mobile data-collection tools for citizen science. In *Proceedings the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing*, San Antonio, TX, USA. DOI: 10.1145/2441776.2441940.

Maisonneuve, N., Stevens, M., & Ochab, B. (2010). Participatory noise pollution monitoring using mobile phones. *Information Policy, 15*, 51-71.

Michelucci, P. (2013). Synthesis and Taxonomy of Human Computation. In P. Michelucci (Ed.), *Handbook of Human Computation* (83-86). New York, NY: Springer.

Mun, M., Reddy, S., Shilton, K., Yau, N., Burke, J., Estrin, D., Hansen, M., Howard, E., West, R., & Boda, P. (2009, June). PEIR, the Personal Environmental Impact Report, as a platform for participatory sensing research. In *Pro-*

*ceedings of the 7th Annual International Conference on Mobile Systems, Applications, and Services*, Krakow, Poland. DOI: 10.1.1.186.8874.

Nissenbaum, H. (2011). *Privacy in Context.* Stanford: Stanford University Press.

Ottinger, G. (2009). Buckets of resistance: Standards and the effectiveness of citizen science. *Science, Technology, & Human Values, 35,*(2), 244-270. DOI: 10.1177/0162243909337121.

Nov, O., Arzay, O., & Anderson, D. (2011, February). Dusting for science: Motivation and participation of digital citizen science volunteers. In *Proceedings of the 2011 iConference*, Seattle, WA, USA. DOI: 10.1145/1940761.1940771.

Paulos, E., Foth, M., Satchell, C., Kim, Y., Dourish, P., & Choi, J. (2008, September). Ubiquitous sustainability: Citizen science and activism. In *Proceedings of the Tenth International Conference on Ubiquitous Computing*, Seoul, South Korea.

Paulos, E., Honicky, R.J., & Hooker, B. (2008). Citizen science: Enabling participatory urbanism. In M. Foth (Ed.), *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City* (414-426). Hershey, PA: IGI Global.

Robson. C., Kandel, S., Heer, J., & Pierce, J. (2011). Data collection by the people, for the people. In *Proceedings of the 2011 ACM Conference on Human Factors in Computing Systems*, Vancouver, BC, Canada.

Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D., & Jacobs, D. (2012). Dynamic changes in motivation in collaborative ecological citizen science projects. In *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work*, Seattle, WA, USA. DOI: 10.1145/2145204.2145238.

Scassa, T., & Chung, H. (2015). *Typology of citizen science projects form an intellectual property perspective*. Woodrow Wilson International Center for Scholars: Washington, DC.

Schwartz, P., & Solove, D. (2014). Reconciling personal information in the United States and European Union. *California Law Review, 877*(102), 877-976.

Sheth, A. (2009). Citizen sensing, social signals, and enriching human experience. *IEEE Internet Computing*, *July-August 2009*, 80-89. DOI: 10.1109/MIC.2009.77.

Shirk, J.,Ballard, H., WIlderman, C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B., Kransy, M. & Bonney, R. (2012). Public participation in scientific research: A framework for deliberate design. *Ecology and Society, 17*(2), 29-49. DOI: 10.5751/ES-04705-170229.

Shilton, K. (2009). Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection. *Communications of the ACM, 52*(11), 48-53. DOI: 10.1145/1592761.1592778.

Shilton, K. (2012). Value levers: Building ethics into design. *Science, Technology, & Human Values, 38*(3), 374-397. DOI: 10.1177/0162243912436985.

Shilton, K., Burke, J., Estrin, D., Govindan, R., & Kang, J. (2009). Designing the personal data stream: Enabling participatory privacy in mobile personal sensing. In *Proceedings of the Research Conference on Communication, Information and Internet Policy*, Washington, DC, USA.

Sullivan, B., Aycrigg, J., Barry, J., Bonney, R., Bruns, N., Cooper, C., Damoulas, T., Dhondt, A., Dietterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J., Fredericks, T., Gebracht, J., Gomes, C., Hochachka, W., Iliff, M., Lagoze, C., La Sorte, F., Merrifield, M., Morris, W., Phillips, T., Reynolds, M., Rodewald, A., Rosenberg, K., Trautmann, N., Wiggins, A., Winkler, D., Wong, W., & Wood, C. (2014). The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation, 169,* 31-40. DOI: 10.1016/j.biocon.2013.11.003.

U.S. Department of Health & Human Services. (1979). *The Belmont Report (*DHEW pub. no. (OS) 78-0012). Washington, DC: National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.

Wallis, J., Mayernik, M., Borgman, C. & Pepe, A. (2010, June). Digital libraries for scientific discovery and reuse: Form vision to practical reality. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, New York, NY, USA. DOI: 10.1145/1816123.1816173.

Wiggins, A. (2013). Free as in puppies: Compensating for ICT Constraints in Citizen Science. In *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing*, San Antonio, TX, USA. DOI: 10.1145/2441776.2441942

Wiggins, A. & Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. In *Proceedings of the 2011 Hawaii International Conference on Systems Science*, Manoa, HI, USA. DOI: 10.1109/HICSS.2011.207.

Wiggins, A., & Crowston, K. (2015). Surveying the citizen science landscape. *First Monday,* 20(1). DOI: 10.5210/fm.v20i1.5520.

Wood, C., Sullivan, B., Iliff, M., Fink, D., & Kelling. S. (2009). eBird: Engaging Birders in Science and Conservation. *PLoS Biology* 9(12), e1001220. DOI: 10.1371/journal.pbio.1001220.

# 9. APPENDIX

| **Project** | **Description** | **Participation model** | **Primary project goals** | **Status** |
|---|---|---|---|---|
| CS1 | Collecting invertebrate observations | Contributory | Investigation | Active |
| CS2 | Invasive species monitoring | Contributory | Conservation | Active |
| CS3 | Plant biodiversity monitoring | Collaborative | Conservation | Active |
| CS4 | Collecting invertebrate observations | Contributory | Investigation | Active |
| CS5 | Plant phenology monitoring | Contributory | Investigation | Active |
| CS6 | Invasive species monitoring | Contributory | Conservation | Active |
| CS7 | Supporting STEM education goals | Co-created | Education | Active |
| CS8 | Collecting weather data | Contributory | Investigation | Active |
| CS9 | Collecting plant observations | Contributory | Investigation | Active |
| CS10 | Collecting invertebrate observations | Contributory | Investigation | Active |
| CS11 | Collecting vertebrate observations | Contributory | Investigation | Active |
| CS12 | Collecting invertebrate observations | Contributory | Investigation | Active |
| CS13 | Collecting invertebrate observations | Contributory | Investigation | Active |
| CS14 | Collecting weather data | Contributory | Investigation | Active |

| | | | | |
|---|---|---|---|---|
| CS15 | Collecting vertebrate observations | Contributory | Conservation | Active |
| PS1 | Sharing seismology data | Contributory | Conservation | Active |
| PS2 | Air quality monitoring | Collegial | Action | Active |
| PS3 | Collecting weather data | Contributory | Investigation | Active |
| PS4 | Measuring cell network connectivity | Contributory | Investigation | Active |
| PS5 | Supporting public health | Contributory | Action | Inactive/ Prototype |
| PS6 | Collecting information on cultural travel routes | Collaborative | Education | Inactive/ Prototype |
| PS7 | Water quality monitoring | Collegial | Action | Inactive/ Prototype |
| PS8 | Supporting public health | Collaborative | Investigation | Inactive/ Prototype |
| PS9 | Supporting national security | Collaborative | Virtual | Active |
| PS10 | Participatory mapping | Collegial | Virtual | Active |
| PS11 | Sharing traffic information | Contributory | Investigation | Inactive/ Prototype |
| PS12 | Volunteered geographic information | Contributory | Action | Active |
| PS13 | Measuring ambient noise | Contributory | Investigation | Active |
| PS14 | Sharing seismology data | Contributory | Conservation | Active |
| PS15 | Sharing seismology data | Contributory | Conservation | Active |

Table 4. Additional information related to theoretical sampling for the 30 projects included in the sample.