Accounting for Privacy in Citizen Science: Ethical Research in a Context of Openness

Anne Bowser
The Wilson Center
Washington, USA
anne.bowser@wilsoncenter.org

Katie Shilton
University of Maryland
College Park, USA
kshilton@umd.edu

Jennifer Preece
University of Maryland
College Park, USA
preece@umd.edu

Elizabeth Warrick
University of Maryland
College Park, USA
ewarrick@umd.edu

ABSTRACT
In citizen science, volunteers collect and share data with researchers, other volunteers, and the public at large. Data shared in citizen science includes information on volunteer location or other sensitive personal information; yet, volunteers do not typically express privacy concerns. This study uses the framework of contextual integrity to understand privacy accounting in the context of citizen science, by analyzing contextual variables including roles; information types; data flows and transmission principles; and, uses, norms, and values. Findings show that uses, norms, and values—including core values shared by researchers and public volunteers, and the motivations of individual volunteers—have a significant impact on privacy accounting. Overall, citizen science volunteers and practitioners share and promote openness and data sharing over protecting privacy. Studying the context of citizen science offers an example of *contextually-appropriate* data sharing that can inform broader questions about research ethics in an age of pervasive data. Based on these findings, this paper offers implications for designing data and information flows and supporting technologies in public and voluntary data sharing projects.

Author Keywords
Citizen science; crowdsourcing; research ethics; privacy; open data; contextual integrity; digital volunteers

ACM Classification Keywords
K.4.1. Computers and Society: Public Policy Issues: Privacy

INTRODUCTION

Citizen science is an increasingly important cluster of activities in which members of the public participate in scientific research to achieve real world goals [2, 17]. The CSCW community has a rich history of promoting and supporting citizen science through research on topics including volunteer motivation [13,37]; newcomer acclimation to online communities [27]; movement through communities and supporting platforms [18]; community-based validation strategies [44]; and the value of information and communication tools and technologies [20, 42]. An additional line of inquiry that unites citizen science practitioners [3, 31] and the broader CSCW community [5, 7, 14, 15, 16] is understanding and protecting privacy in digital contexts that challenge existing research ethics practices and norms.

As an illustrative example of the complexity of privacy and pervasive data collection in citizen science, consider that the project eBird—which asks volunteers to submit checklists of different bird species—has collected over 120 million observations from 150,000 volunteers since 2002 [39]. These data are used in research on topics ranging from species distribution to the spread of infectious disease, and to inform conservation policy and land management decisions. Despite the sophistication of this project, and the value of the data collected, a number of activities may threaten the privacy of citizen science volunteers. Species checklists contain information including exact geo-location; date and time of each checklist; status as "stationary" or "traveling"; duration; party size; and, volunteer name [12]. These checklists may be uploaded via mobile devices in real time, and are accessible by anyone with an Internet connection through eBird's "Explore a Region" feature.

The potential privacy risks generated by such data range from violations of personal autonomy, to algorithmic discrimination enabled by increased tracking [11], to potentially harming personal safety and security [10]. Though research into privacy preferences reveals complicated relationships between privacy concerns and behavior [29], studies increasingly show that adults are wary of the growing trend of tracking personal data using mobile devices, and generally believe that the risks of sharing location information outweigh the benefits [4]. Why, then, are citizen science projects like eBird so successful at recruiting and retaining volunteers?

Some research suggests that incomplete or poorly articulated data policies prevent volunteers from understanding how information is collected and shared [3]. Other scholarship, including Nissenbaum's definition of privacy as *contextual integrity*, suggests that situational variables—including roles, data types, transmission

principles, data uses, and values—influence how people perceive privacy in a particular context [29]. The process that people go through when taking these situational variables into account, and determining the costs and benefits of sharing information, has been referred to as the *privacy calculus* [23]. While privacy calculus emphasizes rational decision-making, other forms of *privacy accounting* embrace a view of the "networked self," in which privacy is understood to be a complex social negotiation involving subjective experience [5]. Drawing on the understanding of contextual integrity as a form of privacy accounting, we identified the following research questions to explore how privacy is understood and accounted for in citizen science:

1. What situational variables—including roles; data types; information flows; data uses; and, norms and values—are important in privacy accounting in citizen science?
2. What are potential threats to volunteer privacy in citizen science, and how are these understood and addressed?
3. Based on privacy accounting in citizen science, how can information flows and interfaces support the privacy needs of volunteers?

Studying these questions in the context of citizen science allows us to understand how privacy and appropriate data sharing operate in this form of pervasive data research. Ethical practices for researchers who wish to collect pervasive data are an ongoing topic of discussion and debate in CSCW [4, 6, 7, 14, 15, 16, 22, 35, 41]. A growing body of empirical work seeks to support this discussion [16]. Understanding privacy accounting in citizen science provides empirical evidence to better understand how forms of participation might impact people's perceptions of privacy in this particular form of data-oriented research. This work benefits the citizen science community by providing an empirical basis for understanding privacy accounting that can be incorporated in the design of projects and supporting technologies. This case study also benefits broader CSCW research by providing examples of contextually appropriate pervasive data sharing to inform larger discussions of research ethics in CSCW.

The remainder of this paper proceeds as follows. The section on "background" discusses contextual privacy and the citizen science domain. Next, the method is described, followed by a presentation of the results. This paper ends with recommendations for designing citizen science data flows and supporting technologies that respect contextual privacy expectations, and a discussion of next steps for advancing ethical pervasive data research.

BACKGROUND

While privacy and research ethics are increasingly raised as important issues in the citizen science community (e.g., [3, 32, 33]) there is a dearth of empirical work exploring these considerations, and an absence of research conducted with citizen science volunteers. One goal of this study is therefore to understand what privacy concerns volunteers and project coordinators actually have, and whether and how these are expressed. With this goal in mind, this section begins with a brief overview of the privacy theory that grounds this work. Following this introduction, key aspects of the citizen science context are described in order to provide background for the study.

Privacy in the Context of Citizen Science

The literature defining privacy is diverse [29, 30, 37]. Privacy has been defined as the right to be left alone, the right to control personal information, a stable attribute of personality, and a culturally-variable social construct [37]. The present study draws upon a growing empirical and theoretical understanding of individuals' expectations about information transmission and use as dependent upon social context [8, 24, 29]. Nissenbaum's foundational theory of *contextual integrity* posits that individuals provide information within a particular social context, and with an understanding of the implicit and explicit information norms that govern that context [29]. The key features of contexts include:

- Who/Roles—people and organizations who are information senders, recipients, and the subjects of the information.
- What/Information—the data types being transmitted.
- How/Information flows and transmission principles—enablers and constraints on the flow of information.
- Why/Uses, norms, and values—the purpose of information collection; and, shared values and norms in a given social context [29].

Empirical work that builds on contextual integrity has explored how individuals negotiate the information norms of various social contexts. Work by Martin and Shilton, for example, measured context-dependent user privacy expectations for mobile applications [24]. Findings demonstrated that very common activities of mobile companies (harvesting and tracking location data, contacts, keywords, name, images and friends) do not meet users' privacy expectations. But these differences are modulated by both data type and social context. For example, consumers expect weather applications to use location data, but do not expect music or banking applications to use location data [40]. Their work supports the notion of "privacy calculus": informal equations engaged in by individuals as they weigh expectations about data collection and use dictated by actors, roles, and contexts.

Individuals who conduct privacy calculus might be pragmatists who exchange information for specific benefits, e.g., better relationships, power, team cohesion, etc. [40]. Or individuals may develop privacy expectations with the costs and benefits of sharing information in mind [19]. For example, volunteers who contribute to eBird might be motivated by the benefits of developing and maintaining

species life lists, receiving rare bird alerts, and/or using eBird's data visualization tools [39]. Privacy accounting can also be understood as less rational and more subjective, having to do with negotiation of boundaries, identity, and the interplay of people and their communities [8]. Privacy research in social networks has frequently focused on this understanding of privacy [e.g. 21, 40]. Under this framework, eBird volunteers might see their contributions as a fundamental part of connecting with a community, sharing knowledge, and contributing to the greater good. To distinguish more subjective notions of privacy decision-making, we refer to this "softer" calculus as *accounting* for privacy.

Key to all contextual definitions of privacy is how these norms work together within a specific context, in this case the context of citizen science. "Contextual integrity" depends on whether shared, situational norms are respected (preserving contextual integrity) or breached (violating contextual integrity). Contextual understandings of privacy have been used to evaluate general policy approaches, for example by suggesting that openness and transparency are necessary but insufficient protections [1]; to examine expectations in specific technological contexts, such as web and mobile applications [24]; and, to understand privacy in specific application domains, such as medicine [29], collaborative work [28], and now citizen science.

Citizen Science

Citizen science is a form of collaboration where members of the public contribute to scientific research [2, 5]. While eBird was introduced as just one example of a citizen science organization (or *project)*, the full diversity of this field is explored below.

Who/Roles

Key roles include researchers conducting citizen science projects (*project coordinators*) and public contributors to citizen science (*volunteers*). Depending on project governance model, project coordinators and volunteers play a number of roles. Three common governance models include: contributory projects, where scientists are responsible for leading research, but solicit data from volunteers; collaborative projects, where scientists involve volunteers in multiple aspects of research, for example both data collection and data analysis; and, co-created projects, where scientists and volunteers work together as partners on numerous aspects of research design [36].

Citizen science projects increasingly encourage expanded models of participation, where a volunteer's role grows beyond contributory data collection to more active engagement in science and policymaking processes [7, 17]. These expanded models challenge traditional roles by allowing volunteers to act in a capacity previously reserved for professional researchers. For example, in collaborative and co-created projects volunteers may contribute to data analysis and interpretation, which requires privileged access to the raw data of other volunteers, including sensitive information.

What/Information

Citizen science involves volunteers in a range of activities. One typology describes the types of information in citizen science through the lens of participation tasks [42]. Common tasks include observation; species identification; classification or tagging; data entry; measurement; specimen/sample collection; geolocation; photography; and, data analysis, among others. These tasks may differ depending on the scientific research domain. For example, a plant phenology project might ask volunteers to collect information including the classification of a local tree; while a crowdsourcing project might ask volunteers to classify the shape of a galaxy. Many projects involve numerous tasks, and different types of information.

How/Information flows and transmission principles

While citizen science is hundreds of years old, the field is experiencing rapid growth facilitated by new wireless, cellular, and satellite technologies. These technologies (which include GPS-enabled smartphones; DIY hardware and software sensors; and shared interfaces such as tabletops [31]) build upon and expand traditional information flows where data are often shared via paper and pencil data sheets [43]. New information flows pose new challenges to volunteer privacy; for example, moving from paper and pencil to mobile means that information about location can be collected and uploaded in real time. Projects might publish, for example, geolocation data in real time, or might enforce a delay to protect participants' locations. In return, projects might guarantee confidentiality by altering data before publication.

Why/Uses, norms, and values

Citizen science approaches the question of *why* activities are conducted from a number of angles. Some researchers study project goals, for example by delineating different types of projects including action-oriented projects, which encourage intervention in local concerns; conservation projects, which support stewardship and natural resource management; and, education projects, which take learning and outreach as primary goals [43]. Other researchers approach the *why* question from another angle, by studying the motivations of citizen science volunteers. For example, Rotman and colleagues found that egoistic motivations, such as personal interest in a topic, drive both initial and sustained participation, while collectivist and altruistic motivations, including community involvement and advocacy, were very important for motivating ongoing participate over time [34]. Motivations for participation are an important topic in the citizen science literature, and a critical part of the shared norms and values of citizen science. Motivations will return as an important theme in our data, as discussed in the Results section.

METHOD

This study is a qualitative exploration of how privacy accounting occurs in the context of citizen science. Drawing on the key roles identified in citizen science, the sampling strategy targeted two populations: project coordinators and volunteers. Project coordinators shared their experiences through individual semi-structured interviews, while volunteers participated in a focus group, as described below. Following data collection, the researchers created a codebook based on Nissenbaum's contextual integrity framework [29], which guided the primary analysis of the interviews and the focus group transcripts. This study was reviewed and approved by the researchers' university Institutional Review Board (IRB).

Participants

Drawing on our understanding of how different topics, research activities, and governance models support a range of citizen science experiences, we used a purposive sampling technique [26] to recruit participants to a study on why volunteers participate in different types of citizen science activities. Of the 13 project coordinators we recruited, 9 ran a dedicated initiative (e.g. a single environmental monitoring or participatory mapping project); the remaining 4 supported more than one (and often numerous) citizen science projects, for example by collaborating with a number of communities around similar monitoring activities, or by providing technical infrastructure to support multiple projects. The projects that coordinators represented came from a wide range of disciplines and scientific fields, including environmental monitoring (n= 4); biodiversity and conservation (n= 3); biology (n= 1); ecology (n= 1); participatory mapping (n= 1); and, public health (n= 1). They also represented a range of governance models [36], and could be characterized as supporting contributory (n= 8), collaborative (n= 3), and co-created (n= 2) activities. These projects involved volunteers in tasks including observation; species identification; classification or tagging; data entry; measurement; sample analysis; site selection; geolocation; photography; and, data analysis [42]. Out of respect for the sensitive information shared by project coordinators and our own IRB protocols, no additional potentially identifying details are described in this paper.

While we initially hoped project coordinators would refer us to their volunteers, many hesitated to broker these connections, either because they did not want to saturate volunteers with requests to participate in research on citizen science or because they did not wish to proactively raise privacy concerns. For this reason we decided to invite volunteers to a focus group held in conjunction with a citizen science networking event at a natural history museum. This allowed us to recruit particularly engaged volunteers familiar with the culture and norms of citizen science. Fourteen volunteers attended the focus group. Each reported experience with multiple citizen science projects; many could not list the exact number of projects they contributed to, or identify each by name. This is consistent with research that suggests that volunteers "dabble" with a number of projects before committing to longer-term participation in a few [13]. At the same time, volunteers did name specific projects during the course of the discussion. This helped the authors conclude that the diversity of projects contributed to by volunteers exceeded the diversity of projects run by coordinators.

Neither project coordinators nor volunteers were financially compensated for participating in this study. Volunteers were offered a casual meal prior to the focus group.

Interview and Focus Group Procedure

Interviews and focus groups followed a semi-structured protocol, where researchers committed to asking a number of established core questions, but allowed for deviation from a formal script to follow-up on interesting points and respect conversational flow. Project coordinators and volunteers were asked variations of the same questions tailored to their roles. The interview and focus group protocols began by establishing history and duration of participation; for example, both project coordinators and volunteers were asked to, *"Please explain your involvement with [citizen science project]. When did you begin working with this project?"* and to *"Explain your level of involvement with other citizen science projects."* Subsequent questions explored perceptions of volunteer participation in greater detail. For example, project coordinators were asked, *"What do you think motivates volunteers to participate in [citizens science project]?"* and *"What kinds of data are collected and analyzed?"* Conversely, volunteers were asked, *"What motivates you to participate in [citizen science projects]?*"

We continued our protocol with the general question, "*Do you have any concerns related to participation in [citizen science project]?"* Follow-up questions designed to elicit privacy concerns were asked on an as-needed basis, and included *"Did any of the data [collected/ analyzed] feel sensitive to you?"* And, finally, *"Do you have any concerns related to privacy?"* By moving from less-leading to more-leading questions, we were able to collect data on our primary area of interest—privacy concerns—while also ensuring that we could analyze how prominent such concerns were to coordinators and volunteers. Following each interview, participants were thanked, debriefed about the primary purpose of this study, and invited to contact the first author with follow-up questions. All interviews and focus groups were audio recorded and later transcribed.

Data Analysis

The first stage of data analysis involved constructing a codebook, which included codes for four key privacy norms supported by Nissenbaum's framework (Who/Roles; What/ Information; How/Transmission principles; and, Why/Uses, norms, and values [29]). Four researchers inductively coded a small portion of the data corpus using Dedoose software with the goal of evaluating and expanding the initial

codebook to more appropriately fit the unique context of data collection. The finalized codebook included the four initial codes, as well as additional codes designating emerging categories of interest, including "privacy concern," "motivation," and "project design." Based on this new codebook, two researchers deductively coded the entire corpus of data.

To begin constructing a cohesive thematic thread around the three research questions, all excerpts marked with the code "privacy concern" were retrieved, and organized through affinity diagramming [26]. During this process, interdependences between the concept of "privacy concern" and other concepts, such as "motivation," began to emerge. In these cases, researchers retrieved additional excerpts associated with these codes to explore the interdependencies more deeply. As analysis began to produce a cohesive narrative, the researchers continually challenged their understanding of the data by deliberately searching for and reconciling conflicting viewpoints through group discussion. Thus, while the first author led the data analysis process, the research team worked in close collaboration to discuss and agree upon the meaning of key concepts and overarching themes. The result of this analysis is presented below.

RESULTS

This section begins by (i) exploring key situational variables that are important in citizen science privacy accounting, before (ii) exploring potential threats to volunteer privacy in citizen science raised by volunteers and coordinators, and (iii) evaluating how potential privacy threats are understood and accounted for. When direct quotations are given, the letters "PC" designate the words of project coordinators, while "V" indicates a quotation from a citizen science volunteer.

Key Situational Variables

According to contextual integrity, privacy concerns should be understood in the context of the implicit and explicit information norms of citizen science [29]. This section draws on the interviews to explore key situational variables important for privacy accounting in citizen science.

Roles

Our study began by abstracting two user groups: project coordinators, who were often professional scientists, and "lay" citizen science volunteers. In reality this distinction is not so clear. Project coordinators can be committed volunteers, as described by PC 8: *"I attend [project] workshops, their field days, their field trainings, their in-classroom trainings and I, myself, am a volunteer for that project."* Volunteers may also be professional scientists. Focus group participants included V 5, a former microbiologist; V 13, a graduate with *"a bachelor's degree in physics"*; and V 3, someone *"involved professionally, at the university level in research."* In general, volunteers are a heterogeneous group, and include families; *"indigenous communities"* (PC 10);*"retirees who aren't ready to just let everything go and lay on the beach"* (PC 5)*; "mid-life career changers… just looking for something different to do"* (PC 5); *"college students…helping to build their resume out"* (PC 5); and elementary- or middle-school *"teachers and their students"* (PC 13).

Many citizen science projects are understaffed and/or underfunded [42]. Institutional resources to support project coordinators vary; a few coordinators we interviewed were, as PC 5 put it, *"doing this for free… I've never been paid to do it ever."* As a result, coordinators of underfunded projects often play a number of roles. Project Coordinator 7, formally trained in informal science education, was encouraged to "*teach yourself a little HTML"* rather than hire or contract a professional app developer. Teams that are spread too thin also adopt non-optimal workflows. For example, while one project would like to review public comments from volunteers on a daily basis, they compromise on weekly review due to staff time constraints.

But other projects, such as those supported by or run out of government agencies, enjoy access to resources such as legal teams. Such access was often mentioned in the context of complying with legal privacy mandates such as the Children's Online Privacy Protection Act, or COPPA. As Coordinator 6 explained, *"We were an organization of about 1,200 people… We were a small group of people within a group, trying to do a citizen science program, but our entire website, everything we asked was reviewed in part, because of the COPPA laws*." Recognizing both the burden of legal compliance and the privilege of a legal team to facilitate compliance, Coordinator 6 continued, *"I could certainly understand …if you were a small project starting up, didn't have a national center or attorneys on staff, yeah, you might do things that compromise privacy, not because you were a horrible person, but because you just didn't think about those things."*

As noted earlier, citizen science projects have different governance models, allowing volunteers to play different roles in the research process [36]. In contributory projects, the main responsibility for volunteers is to *"upload their data"* (PC 2) by following set protocols. Project coordinators in co-created projects partner with volunteers to determine key aspects of project design, including sampling (*"It's up to the community whether they want to use their own kit in a residential area, or they want to go to a school…"* PC 9) and data storage and access permissions. For example, Coordinator 10 is exploring the possibility of *"having a public key that sits with the [community], where every time you want to access the data, you need them to open the data for you."*

Participants also identified stakeholders not directly involved in citizen science activities. These include other projects and scientists conducting research through other means. Government agencies also use citizen science data, either for regulatory enforcement or forecasting. Finally, projects that go through research ethics reviews, like those

sponsored by university IRBs, must contend with staffers and researchers on review boards.

Information and data

Our interview participants understood citizen science data as facts or other information collected, analyzed, and used during the citizen science research process. Broad types of data included environmental monitoring data (e.g., on temperature or water quality); non-human species observations (e.g., avian presence, absence, or count); phenological observations (e.g., reporting the current state of a species); human biological data (e.g., urine samples); and, data contained in geographic information systems (e.g., including data about both natural and man-made structures). Consistent with [35], data take the form of written measurements or observations (including close-ended observations, such as checkboxes, and open-ended observations, including comments); images or photographs; audio recordings; video recordings; direct samples (e.g., of human biological data or an invasive species); and, geographic location (e.g., including exact latitude and longitudinal coordinates, or mailing address).

Data are accompanied by metadata, or information that documents and adds value to primary data. Common metadata include volunteer name or username (either assigned by the project, or selected by each volunteer); contact information in the form of email address and/or phone number; volunteer location, captured as GIS coordinates and/or IP address; the data and time an activity took place; and images or audio recordings. In addition to data and metadata, projects collect additional information to facilitate volunteer management. Such information may include email addresses; mailing addresses; telephone number; and social media data, such as Facebook usernames and Twitter handles. Some projects also collect *"an optional demographic survey*" (PC 6).

Information flows and transmission principles

Citizen science is predicated on a novel information flow: the direct exchange of data, analysis, and other information between professional scientists and public volunteers. In some cases, information is shared through paper-and-pencil forms. But information more frequently flows through browsers and apps accessed on laptops; desktop computers; and portable devices such as Smartphones and networked sensors. Some projects offer multiple options for data submission. In these cases, volunteers select the information flow with which they feel most comfortable. As Volunteer 14 put it: *"I've never used a Smartphone option for myself and many others haven't…it's not necessary."*

In simple information flows typical of many contributory projects [36], volunteers consent to share data primarily or exclusively with a project coordinator for direct use in research. In other projects and governance models, volunteers also consent to share data with other volunteers, including direct collaborators (e.g., when citizen science is conducted in the context of formal education); with other, unknown volunteers (e.g., when all registered volunteers enjoy privileged access to project data); or, with third parties (at each volunteer's individual discretion). In such cases, many citizen science project coordinators establish a principle of reciprocity by making all contributed data openly available. As Project Coordinator 11 describes, *"you can download the whole dataset straight from the website, free of charge and with no restrictions to it*."

Despite this principle of reciprocity, some transmission constraints may be placed on citizen science data, such as when publication is delayed by a short period of time, typically a single day, to avoid real-time identification of volunteers' locations. Transmission constraints may also be caused by technical limitations, as described by Coordinator 12: *"[Volunteers] take real-time observations and then as soon as they come back into the area near the visitor's center, which does have phone reception and WiFi, the data is immediately uploaded to our servers*."

In long-term projects, information flows and transmission principles change over time, particularly as the technological affordances available to a project change. As Coordinator 3 explained, *"In the past it was very easy for an observer to basically blend into the landscape… but nowadays with GPS…we'd have to identify the location of the station, and we do that, with GPS, as close as the technology will allow."* Coordinator 12 relayed the same challenge: "*The [new] watershed … truthfully, most of it has better cellphone reception. So if we deploy the same technology here, I think there will be more concerns about volunteer participation and a real time access to location data"* (PC 12). New technological affordances for information flows may mean that coordinators need to reconsider whether they should implement transmission principles such as improved data anonymization.

Data uses, norms, and values

Interview and focus group participants discussed values and norms relevant to the context of citizen science, and the motivations of citizen science volunteers. Researchers consider citizen science a valued mechanism for opening traditionally closed science and policy systems to greater public participation [8, 17]. Many participants report that citizen science is dominated by an ethic of openness: *"I think the whole attitude towards it is open source. We're doing this to share it with anyone and everyone so it can be used to benefit bird conservation in any way"* (PC 11). The value of openness is expressed in terms of participation by a range of groups, including *"younger folks, less well-off folks, and various socioeconomic groups that don't participate in citizen science enough"* (PC 12). Openness is also discussed in terms of data. Volunteers value open data, and even find bragging rights in broad information dissemination: *"I'd like to know if my data is being used by other projects. In fact I'd tell my wife and kids*." (V 1).

For many volunteers, personal motivations for sharing data outweigh the risk associated with ceding their privacy. Motivations include attribution: *"Most people are actually quite chuffed to see their name on there and knowing that the world can see what they're doing and making a difference"* (PC 11). Personal interests are a second important motivation, especially at the beginning of participation. As one volunteer reports: "*At the time I was really into paleontology. And I started looking for paleo projects"* (V 6). Volunteers are also motivated by a more general *"Curiosity. You just want to know things, like the wildlife around your home"* (V 6).

Many volunteers like to be *"outside in nature"* (V 7). For some, participation is most valuable when connected to a local environment, for example when students "*connect to what they've seen in their backyard"* (V 2). Participation can also help volunteers explore new local areas, especially if *"They wouldn't have chosen necessarily to hike there recreationally themselves if it weren't for their scientific contributions"* (PC 8). But locality is not a necessary condition for all: Volunteer 9, who participates in national-scale projects, believes that *"for kids it's probably the immediacy of the local part,"* but adds, *"for me as an adult it doesn't really matter too much. So I guess it just depends on the person."*

Locality can translate to direct personal connection, for example when science is a collaboration with *"Dr. Robert, from upstairs"* (V 8). Place-based projects also support civic connection to a local community, especially when volunteers access local data: *"You can say, 'hey, I want to see…all the data collected in [county] for last year. And you start seeing patterns. So you can start thinking: my goodness, and my stuff is in there. And I own part of that now"* (V 4). One additional important motivation is socialization: *"You've got friends, you go out with other people and it's really like a party"* (PC 7).

Potential Threats to Volunteer Privacy

Data collection, however well-intentioned, was not all openness and socialization. Both volunteers and project coordinators understood potential privacy threats associated with participation in citizen science.

Location privacy

Almost all projects collect information on a volunteer's location during a registration process. As PC 13 put it: *"If you're in our database now, we could find you."* And when volunteers submit a geo-referenced observation, they simultaneously re-identify their location with each data point submitted. For many projects, observation location is the single most important piece of metadata: "*In order to make the data of any value, it has to be known where the data's coming from"* (PC 4). Projects that share public maps of citizen science data recognize that *"anybody can go and click on that data point…if you have sampled at your house, your house shows up"* (V 6). This may be a source of discomfort for those who consider privacy as the right to be left alone. Charting home addresses can also present security issues, for example when *"someone will inadvertently put a comment to say, 'temperature was 79 degrees, and by the way this is my last report for the next week because I'm going out of town."* (PC 3).

The link between observation and volunteer location is especially problematic when data is shared in *"real time"* (PC 8; PC 12; V 1): *"Many of the areas on the water are remote-ish, and if you're far away from a road or a vehicle or other people than broadcasting exactly where you are leads to a number of personal safety concerns"* (PC 12). Location privacy concerns are exacerbated when volunteers come from vulnerable populations, including *"illegal migrants"* (PC 10) and *"kids"* (PC 13; V 9), or when activities are done *"repeatedly and predictably"* (PC 12), thus supporting inferences about routines. Adding to issues of human privacy are concerns about sharing *"threatened and endangered species' locations"* (V 11) and *"potentially opening [species] up to poaching"* (PC 8).

Personal information

Projects collect, and sometimes disclose, personal information about volunteers. As one project coordinator explains, *"we have on our website a community attribution page, and every person who has signed up to participate is listed with their city. So for example, you would be listed as [Lauren N., Virginia, USA]…we ask that if our data is used, that people look at this and they link to that page because indeed, it's a group effort"* (PC 6). Volunteers also share their own personal information in ways unplanned or unpredicted by project coordinators. One coordinator notes, *"sometimes, in the comments area, people will unknowingly actually self-identify and they will give us an address or they will give some personal information."* (PC 11).

Triangulating location and personal information

As illustrated by the example of eBird, it is often possible to retrieve data documenting observer location alongside personal information like full names. Triangulation also happens on other platforms, for example when volunteers share information in a Facebook group. As one project coordinator shared, "*When someone will mention something…four other people will pipe in, 'well post your [identification] number, please. If you're going to post something on here, we want to know your [identification] number too'….and most people are happy to say, 'I am [486221611].'"* (PC 1).

Other potential privacy threats

Volunteers often submit data with the expectation that it will be used for a specific purpose, such as answering a scientific research question or informing species management. But citizen science data collected for one purpose is often re-used in other contexts, including other research projects, or on social media for communication or promotion. While parameters for acceptable re-use are sometimes documented in data policies, in other cases projects may share data in new ways without altering

policies or informing volunteers. Or, projects may contact volunteers with questions regarding reuse: *"I haven't had anyone sign a photo release form or anything else…[so] if I'm going to post something like that I always ask everybody, are you OK with this going up on Facebook"* (PC 7). Contacting volunteers also has implications for the understanding of privacy as the right to be left alone. Some project coordinators do note, *"sometimes they get tired of hearing from us as is… and [a second organization] can… with our policy, feel free to contact by phone, that instantaneously, any observer."* (PC 3).

Awareness of Privacy in Citizen Science

Volunteers become aware of privacy concerns in a number of ways. Ideally, information on key facets of participation—particularly the types of data collected, and relevant information flows—is posted to a project's website. But data policies are often opaque, or insufficiently documented [3]. Noting the similarities between two different citizen science projects, one volunteer wondered, "*should I do both of them? Should I only do the one?... do they share their data, or is it the same project…I do believe they share, but I don't know that*" (V 4). In addition, when focus group participants were asked whether they typically read citizen science data policies, not a single one answered affirmatively. One volunteer *"can't remember, I must have flipped through them* (V2)*."* Another, whose children also participate in citizen science through classroom-based education, offers: *"I'm not the teacher. I'm the parent. So, I would do it for my own kids, for us"* (V 1). This finding is supported by broader research that suggests policies are neither a clear nor a comprehensive way to convey privacy concerns or data and information flows to users [25].

Instead of reading data policies, some volunteers learn about information flows by experimenting with data on project websites, or *"testing out the limits of what you can access"* (V 2). Awareness of information flows often comes from seeing data from other volunteers: *"Where I was looking at my neighborhood just to see what other people were seeing ...I found this one guy where I can track every single place he goes. And his house. I know that he lives in this house, and he goes to the same nature park after work practically every day."* (V 4). Volunteers also learn about information flows by seeing their own data. One recalls: *"seeing my address, my house… I knew that it was our house. We had just submitted some [data]. And I remember it was like 'oh cool that's ours.'"* (V 1).

Finally, project coordinators educate volunteers about the privacy implications of their actions on an as-needed basis. One explains, *"we write to the volunteer and say, 'hey we took out the wording in your report but in the future don't announce that you're leaving your house'"* (PC 3).

Accounting for Privacy in Citizen Science

Despite the issues described above, and general awareness of potential threats to privacy, the majority of project coordinators reported that volunteers do not typically raise privacy concerns during participation. As PC 11 put it: *"We've not heard anyone express these concerns, and I think we're quite confident on that."* PC 12 agreed: *"I've been very keen on sharing and connecting with people, and [they] have not thought about privacy at all."* PC 6 relayed: *"We have never had anybody express concerns, because I think we are and have always been fairly conservative in the information we share."*

A few project coordinators shared exceptions. Coordinator 8 must contend with differing individual privacy preferences: *"We have quite a few people who want to share who they are, where their location is, compared to others who are really quite scared…actually it's a challenge to meet the needs of those who share their information with those who really want to keep it private."* Coordinator 2 sometimes receives requests for location obfuscation or fuzzing: *"we've had volunteers ask to put [their location] more in the street, maybe somewhat of a guess for which of four or five houses it might be… we do it."* Project coordinators also reported that stakeholders outside the immediate citizen science community, including other researchers and IRBs, considered their work problematic from a privacy perspective: *"The community themselves were not worried about that at all…it's actually academics only that were a problem"* (PC 9).

During the focus group volunteers did not voice privacy concerns prior to prompting from the authors, even when asked about general participation concerns, suggesting that privacy concerns did not rise to the level of primary concerns. When asked directly about privacy, the majority understood the implications of different activities, but still did not express concerns. Some volunteers pointed to general social norms around privacy. Most agreed, *"In today's day and age, you can't say 'I'm off the web, I'm not participating, nobody knows where I am"* (V 4). Many volunteers believed they had already lost the struggle for privacy: *"Everyone in the room is going to say I don't want you to know where I live, my income, these are private things. But again they're just out there"* (V 1). Volunteers also believed that while extremely privacy-conscious individuals exist, these people are *"not signing up for citizen science projects where you send data in"* (V 1). In other words, citizen science participants may be a self-selecting group already willing to take *"a calculated risk*" (V 9) by sharing private information.

But in addition to expressing consistency with general social norms, project coordinators and volunteers identified a number of contextually-specific reasons why "calculated risks" are worth taking with their data. These are related to the variables outlined in Nissenbaum's framework [29], particularly the motivations of citizen science volunteers. For example, many volunteers report *"trust with scientists-you give your email, you trust that they're not going to sell it"* (V 2). Scientists are considered trusted experts, and sharing data with scientists aligns with Nissenbaum's

understanding of appropriate *roles* for data sharing. Volunteers may also place their trust in individuals they know: including, *"Dr. Robert, from upstairs"* (V 8).

Attribution and communication are additional motivations for participation that encourage norms of openness over privacy: "*When I do get feedback… that helps keep me involved in that project. Because somebody is paying attention to me."* Volunteers also appreciated the ways that open information flows encouraged learning. For example, having access to raw data *"might spark some other question towards the project, that I didn't know about… or my kids didn't know about. And then they'd be like, 'well what was there?' or 'why is that in that location?' It would just be curiosity sparking."* (V 7). Volunteers view open information flows as appropriate because full access to data matches their expectations of citizen science as a source of learning.

The values and norms of citizen science explicitly promote data sharing to achieve a greater good. Many volunteers are motivated *"to see the data used to improve environmental monitoring, or create more stringency on regulation. Or cleanup"* (PC 9). The desire to be useful may stem from a collectivist or altruistic mentality, or alternatively personal pride: *"being able to go on a site and see your data point based on your address…is gratifying"* (V 4). Volunteers also enjoy seeing *"how [your contribution] relates to other people's data."* This data transparency can help volunteers find each other, which draws on social motivations for participation: *"I remember I got the impression, hey there's someone down the street that's interested in this too, that's cool…most of the time if I run into somebody you know, we tell them everything we're doing anyway. And invite them to come do it with us"* (V 1). Interestingly, while a few project coordinators suggested that privacy includes the right to be left alone, not a single volunteer raised this point. As these quotations suggest, volunteers aren't joining citizen science projects to be left alone: because socialization and community inclusion are such important motivations, the norms of participation are communal.

Despite a general emphasis on openness and data sharing, participants did express concerns over sharing certain types of data**.** The hardest line was drawn around sharing information about children: "*you shouldn't share a child's name, bottom line"* (V 8). Some volunteers prefer location to be captured in obfuscated formats. One appreciated a project that *"doesn't tell you the address, it just shows you the location…and I'm happy with that, I'm happy with that being out there"* (V 4). Volunteers are also more willing to share location information in some areas than others: *"On our campus we have beautiful woodlands, but it is a private piece of property and they would not want us to open it up and invite people to come look at whatever"* (V 10). In addition, some information flows are preferable to others. *"[If] you're talking about an app, that you take with you everywhere, and you're observing all kinds of stuff…verses participating in a project that is collecting data, maybe you do the project, you do the data collection, you send it in, you're done…I think it's totally different issues"* (V 2). Single geo-referenced observations inspired much less concern than streaming data or constant location tracking.

DISCUSSION

This research provides empirical evidence for theories of privacy as contextual integrity, which posit that privacy concerns are not primarily individual traits but rather based upon expectations in social contexts [29, 30, 37]. Specifically, this study shows that citizen science volunteers take complex social factors and norms *into account* while making judgments about when and how to be concerned about privacy in citizen science. Both project coordinators and volunteers are aware of potential privacy violations associated with participation in citizen science. Yet, the norms and values of this context promote a shared culture that prioritizes openness, rather than data protection.

Volunteers trust project leaders, are motivated by their relationships with leaders, and are motivated to share and socialize with other volunteers. Volunteers appreciate the enhanced access to knowledge that comes with open data. Additionally, volunteers are motivated by the positive social, scientific, and environmental outcomes of citizen science, and inspired to see their data used to further these goals. For all these reasons, volunteers enter projects willing and ready to share personal data. The relative absence of privacy concerns from citizen science contexts, despite pervasive data collection that might alarm individuals in other contexts, indicates that the values and norms of citizen science make data collection a perception of *sharing and contribution* within a broad community of volunteers, project coordinators, and the interested public, rather than a perception of *taking*. Understanding these norms can help citizen science researchers and practitioners design projects and supporting technologies based on grounded, contextually specific privacy expectations [16].

Protecting Privacy in Citizen Science

The findings from this research do not mean that citizen science projects do not need to worry about privacy. Earning the trust that volunteers place in these projects can be challenging for overworked and under-resourced teams. This section discusses common technical, workflow and policy changes that projects can adopt to ensure that such trust is well-founded. To best respect contextual privacy expectations, such changes should be implemented in ways that respect the unique features of a project and the motivations of that project's volunteers

Project coordinators use a variety of mechanisms, including technological safeguards and data policies, to protect volunteer privacy. One common technique is location fuzzing or cloaking [39]. As one coordinator describes, "*Some systems have what they call 'fuzzy data' and that means that they desensitize or de-specify the exact location of a threatened and endangered species…and if you*

reported it, you would know…if you're the project coordinator, you would know, but nobody else would know" (PC 8). Notably, while obfuscation may be designed to protect a non-human species, some volunteers do note human benefits to this feature (e.g., *"to give people the idea, well this is the area, without saying come on over to my house*" (V 3)).

A second privacy protection technique is restricting the amount of personal information collected and shared. As knowing who submits data is scientifically important, most projects associate a volunteer's name or username with their data. Regarding the decision to publish such information, one project coordinator notes, *"We assumed that mostly people don't want to. We don't start from the assumption that people want to, we start from the assumption that they don't"* (PC 10). Project coordinators may assign each volunteer an anonymous identification, or allow volunteers to create their own username. This later approach may be problematic: *"The adult volunteer community we work with…might use the same login for their bank and they might use the same login for all their other accounts…so in some ways it could be argued that it's even more risky"* (PC 8). Thus, mechanisms for restricting personal information are necessarily project-specific.

Other citizen science projects restrict participation to certain populations, for example by excluding children: *"If you were to want to register to be a [participant], there's a box you have to check that says, 'I am over 13'"* (PC 6). Projects targeting school groups may also ask teachers to register and upload data, rather than allowing students to participate directly. Finally, many projects support the privacy preferences of volunteers through well-documented data policies that allow people to make meaningful choices about whether and how to participate [3].

Implications for Data Flows & Technology Design

It is clear that volunteers idealize citizen science projects as open systems, and reward openness with sharing and contribution. Basic privacy precautions such as data obfuscation and minimizing personal information collection are excellent first steps for projects that hope to retain trust and protect volunteer data. Both of these solutions are cheap and scalable best practices for privacy protection.

Designing flexible data flows can further improve the relationship between projects and volunteers. For example, some projects provide different modalities for data submission (paper and pencil; computer; smartphone in the field), allowing volunteers to weigh the costs and benefits of different ways of participating. Citizen science projects might also allow volunteers to change their location preferences in various situations, for example by allowing volunteers to alternately share a street address or drop a pin on a map. These forms of flexibility can be embedded in project data flows and technology design. In addition, flexibility can be considered as a core value when citizen science projects invite volunteers to co-design data collection and sharing protocols, or prototype and develop supporting technologies through participatory design.

Projects that deal with very sensitive data, whether health data or data on a threatened or endangered species, might consider additional ways to support privacy by design. Advanced forms of notice can be built into data collection apps. Designing technologies to remind volunteers about the parameters of participation as they unfold can also relieve volunteers from the burden of having to read and understand complex data policies. More sophisticated approaches might involve filtering data as it is submitted. Just as eBird checks for location [39] to see if data about a particular bird sighting is feasible (e.g., are there really parrots in Maryland), projects could check for unintentionally revealing patterns of behavior. For example, if a user visits a remote farm location at the same time each day or week, a notification might prompt her to consider whether this behavior is advisable, and suggest small ways of tweaking data flows. Flexibility and notice can help projects avoid restricting participation to certain populations, and provide measures to support control-oriented approaches to privacy [37].

An additional recommendation may be drawn from the example of volunteers who learned about potential privacy concerns by experimenting with the data tools provided by the project. Project training processes should incorporate substantial time for volunteers to understand the project's data flows through use of both data collection and aggregation systems in the contexts where these activities unfold. Training might also incorporate brief modules on safe privacy practices, which would explain the options for participating that a project supports.

Privacy and Research Ethics in CSCW

Finally, this research can contribute to larger discussions of privacy and ethical considerations in CSCW research. Social media use echoes citizen science as an area in which privacy norms are impacted by the value of participation, as users account for reduced interpersonal privacy because they understand the norms and benefits of sharing [5, 22, 38]. For this reason, the contextual expectations of practitioners and volunteers in citizen science research may align with the expectation of participants in other contextually bounded contexts. In line with our research, Brown et al. note that the social benefits of participation in research in contexts including social media platforms like Instagram may be "hampered" by anonymity, and argue that the need for acknowledgement in the co-creation of research often outweighs the desire for privacy [7].

On the other hand, social media *research* is frequently hampered by the opposite effect: only occasionally do research projects in social computing adopt the participatory affordances of citizen science. In contrast researchers who enter chat rooms to study (rather than participate in) online communities violate contextual privacy expectations and may be forcibly removed [15, 16].

Empirical examination of ethics questions in contrasting research contexts provides a valuable starting point for understanding how, when, and why communities and individuals may value and promote openness and sharing over privacy. The example of citizen science suggests that norms may skew towards openness when researchers and participants are seen as members of the same community with similar goals and shared values. Our findings also suggest that adding participatory affordances might mitigate many of the privacy concerns currently expressed by publics who object to data scraping or unknowing experimentation with their social media data. A model where volunteers might *donate* their social media data for research, for example, could advance a form of "citizen" social computing research. Public contribution to social computing research could shift power balances to enable more open participant privacy expectations and also, building off the general learning gains associated with citizen science [2], potentially lead to enhanced data literacy as public awareness of social media research grows. Additional empirical research in diverse pervasive data sharing contexts can benefit each context individually while also facilitating comparisons to promote generalized ethics principles [16]. In this way, the CSCW community can use empirical research as a stepping stone towards a shared understanding of contextually-specific research norms.

These findings also present important implications for advancing discussions on ethical oversight [7, 16, 35]. Current research regulations (for example, enforcement of the Common Rule in the U.S. by Institutional Review Boards) seldom take into account the different ethical calculus of various research practices. Furthermore, because many granting agencies require IRB approval *before* a project is funded and volunteers are recruited, it can be difficult for citizen science practitioners to determine ethical best practices in coordination with volunteers. Yet our findings suggest that, as many privacy concerns are mitigated by the trust and shared culture fostered by citizen science models, research regulations should support public consultation in pervasive data research. This might be achieved by adding incentives (or simply granting permissions) for investigators to modify their protocols following community consultation. In addition, organizations like the U.S. Citizen Science Association (CSA) should consider supporting community-based ethical codes of conduct and/or community review processes to better understand and uphold shared contextual norms.

LIMITATIONS & FUTURE WORK

This research focused on small purposive sample of 13 citizen science project coordinators and 14 experienced citizen science volunteers to elicit and begin to answer key questions about privacy accounting in citizen science. Further work is now needed to better understand privacy accounting in citizen science, and to advance conversations about ethics in pervasive data sharing research.

The purposive sampling technique allowed us to reach projects with a wide range of governance models and scientific research tasks [36, 43], in fields including biodiversity and conservation; biology; ecology; participatory mapping; and, public health. However, because most of these projects involved location sharing, the majority of our discussions revolved around location-based privacy concerns. During the process of recruiting interviewees and analyzing our data, we concluded that theoretical saturation was reached in regard to location privacy. More research is needed to move beyond location-based privacy concerns, especially given the growth of genomic-based citizen science projects, which do not always involve a location component yet raise significant privacy concerns [33].

Convening general citizen science volunteers in a focus group allowed us to learn about experiences in a range of projects beyond those represented by the project coordinator sample. However, focus groups may occasionally lead to groupthink. During our analysis we searched for (and found) instances of disagreement, for example around the value of locality; these gave us confidence that groupthink was not a significant issue in this study. Still, interviewing volunteers during future research would ensure that each participant could express potentially significant privacy concerns.

While recruiting experienced volunteers allowed us to collect data about the relationship between norms and values and privacy concerns in citizen science, our sample excludes volunteers who left citizen science projects because of privacy or other concerns. Future work focused on volunteers marginal to, or excluded from, citizen science could provide valuable contrasting data to this study. Future work could also move beyond citizen science to more broadly examine privacy accounting in different types of scientific research.

Each of the project coordinators we interviewed had clearly spent significant time thinking about and/or discussing privacy in their unique projects; it was less clear that these participants considered privacy risks in projects of different types. Shared conversations within the citizen science community, whether through focus groups, professional meetings, or conferences, are required for project coordinators to share experiences and reach a common understanding of privacy accounting and appropriate information flows.

CONCLUSION

Accounting for privacy in citizen science requires accounting for the unique context of these participatory projects. While privacy concerns in this domain are real, they are hardly dominant among volunteers; instead, the context primes volunteers to focus on openness, sharing, and the personal and collective benefits that motivate and accompany participation. In other words, project coordinators and other researchers should understand that in

general, citizen science information flows are contextually appropriate. At the same time, citizen science project coordinators must be mindful of this priming, because volunteers may not raise privacy issues on their own. Instead, privacy should be treated as any other data flow consideration in citizen science: as an opportunity to promote inclusion and autonomy through creative participation and flexible design to understand and support project-specific or situational needs (e.g., through location cloaking). Moving beyond the citizen science and broader scientific community, other pervasive data researchers can learn from the ways that the context of citizen science mitigates participant concerns about privacy and consent. This research is also valuable for contributing to broader agenda setting and discussions around ethical research and practice in locative media and social media contexts.

ACKNOWLEDGEMENTS

We thank the participants in our study, as well as Caren Cooper and Darlene Cavalier. This work was funded by NSF Award SES-1450625.

REFERENCES

1. Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221, 509-514.
2. Rick Bonney, Caren Cooper, Janice Dickinson, Steve Kelling, Tina Phillips, Kenneth Rosenberg, and Jennifer Shirk. 2009. Citizen science: A developing tool for expanding science knowledge and scientific literacy. *Bioscience* 59, 11, 977–84.
3. Anne Bowser and Andrea Wiggins. 2015. Privacy in participatory research: Advancing policy to support Human Computation. *Human Computation* 2, 1, 19-44.
4. danah boyd and Kate Crawford. 2012. Critical questions for big data. *Information, Communication, and Society* 15, 5, 662-679.
5. danah boyd and Alice E. Marwick. 2011. Social privacy in networked publics: Teens' attitudes, practices, and strategies. In *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*. Oxford, UK: SSRN. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1925128]
6. Stacy Branham, Anja Thieme, Lisa Nathan, Steve Harrison, Deborah Tatar, and Patrick Olivier. 2014. Co-creating & Identity-making in CSCW: Revisiting ethics in design research. *In Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW 2014), 305–308. http://dx.doi.org/10.1145/2556420.2558859
7. Barry Brown, Alexandra Weilenmann, Donald McMillan, and Airi Lampinen. 2016. Five provocations for ethical HCI research. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 852-863. http://dx.doi.org/10.1145/2858036.2858313
8. Luigi Ceccaroni, Anne Bowser, and Peter Brenton. 2016. Civic education and citizen science: Definitions, categories, knowledge representation. In Luigi Ceccaroni and Jaume Piera (Eds.), *Analyzing the Role of Citizen Science in Modern Research*. IGI Global.
9. Julie Cohen. 2012. *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice*. Yale University Press.
10. Lorrie Faith Cranor, Patrick Kelley, Norman Sadeh, and Janice Tsai. 2010. Location-sharing technologies: Privacy risks and controls. *Journal of Law and Policy for the Information Society* 6, 2, 1-26.
11. Cynthia Dwork and Deirdre Mulligan. 2013. It's not privacy, and it's not fair. *Stanford Law Review Online* 66.35, 35-40.
12. eBird. Explore Data. 2016. Retrieved May 20, 2016, from http://ebird.org/ebird/eBirdReports?cmd=Start
13. Alexandra Eveleigh, Charlene Jennett, Ann Blandford, Philip Brohan, and Anna L. Cox. 2014. Designing for dabblers and deterring drop-outs in citizen science. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 2985-2994, http://dl.acm.org/citation.cfm?doid=2556288.2557262.
14. Casey Fiesler, Alyson Young, Tamara Peyton, Amy S. Bruckman, Mary Gray, Jeff Hancock, and Wayne Lutters. 2015. Ethics for studying online sociotechnical systems in a big data world. In *Proceedings of the ACM Conference Companion on Computer Supported Cooperative Work & Social Computing* (CSCW 2015), 289–292. http://dx.doi.org/10.1145/2685553.2685558
15. James Hudson and Amy Bruckman. 2005a. "Go away": Participant objections to being studied and the ethics of chatroom research. *The Information Society, 20,* 2, 127-139.
16. James Hudson and Amy Bruckman. 2005b. Using empirical data to reason about Internet research ethics. In H. Gellersen, K. Schmidt, M. Beaudouin-Lafon, & W. Mackay (Eds.), *ECSCW 2005* (pp. 287–306). Springer Netherlands.
17. Alan Irwin. 1995. *Citizen Science: A Study of People, Expertise, and Sustainable Development*. Routledge.
18. Corey Jackson, Carsten Østerlund, Veronica Maidel, Kevin Crowston, and Gabriel Mugar. 2016. Which way did they go? Newcomer movement *through* Zooniverse. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16), 624-635. http://dx.doi.org /10.1145/2818048.2835197
19. Guillermina Jasso. 2006. Factorial survey methods for studying beliefs and judgements. *Sociological Methods & Research* 34, 3, 334-423.

20. Sunyoung Kim, Jennifer Mankoff, and Eric Paulos. 2013. Sensr: Evaluating a flexible framework for authoring mobile data-collection tools for citizen science. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (CSCW '13), 1453-1462. http://dx.doi.org/10.1145/2441776.2441940
21. Han Li, Rathindra Sarathy, and Heng Xu. 2010. Understanding situational online information disclosure as a privacy calculus. *Journal of Computer Information Systems* 51, 1, 62.
22. Heather Richter Lipford, Pamela J. Wisniewski, Cliff Lampe, Lorraine Kisselburgh, and Kelly Caine. 2012. Reconciling privacy with social media. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion* (CSCW '12), 19-20. http://dx.doi.org/10.1145/2141512.2141523
23. Kirstin Martin. 2013. Transaction costs, privacy, and trust: The laudable goals and ultimate failure of notice and choice to respect privacy online. *First Monday* 18, 12, online.
24. Kirsten Martin and Katie Shilton. 2016. Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices. *The Information Society* 32, 3, 200-216.
25. Aleecia McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society* 4, 543, 565.
26. Matthew Miles and Michael Huberman. 1994. *Qualitative Data Analysis: An Expanded Sourcebook.* SAGE Publications, Inc.
27. Gabriel Mugar, Carsten Østerlund, Katie DeVries Hassman, Kevin Crowston, and Corey Brian Jackson. 2014. Planet hunters and seafloor explorers: Legitimate peripheral participation through practice proxies in online citizen science. In *Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & Social Computing* (CSCW '14), 109-119. http://dx.doi.org /10.1145/2531602.2531721
28. Alison Murphy, Madhu Reddy, and Heng Xu. 2014. Privacy practices in collaborative environments: A study of emergency department staff. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '14), 269–282. http://dx.doi.org/10.1145/2531602.2531643
29. Helen Nissenbaum. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford Law Books.
30. Xinru Page, Karen Tang, Fred Stutzman, and Airi Lampinen. 2013. Measuring networked social privacy. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion* (CSCW '13), 315–320. http://dx.doi.org/10.1145/2441955.2442032
31. Jennifer Preece. 2016. Citizen science: New research challenges in HCI. *Int'l J. of Human-Computer Interaction* 32, 8, 585-612.
32. David Resnik, Kevin Elliott, and Aubrey K. Miller. 2015. A framework for addressing ethical issues in citizen science. *Environmental Science & Policy* 54, 475-481.
33. Mark Rothstein, John T. Wilibanks, and Kyle B. Brothers. 2015. Citizen Science on your smartphone: An ELSI research agenda. *Journal of Law, Medicine, & Ethics,* 43.4, 898-903.
34. Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (CSCW '12), 217-226. http://dx.doi.org/10.1145/2145204.2145238
35. Katie Shilton and Sheridan Sayles. 2016. "We aren't all going to be on the same page about ethics": Ethical practices and challenges in research on digital and social media. In *Proceedings of the 2016 49th Hawaii International Conference on System Sciences* (HICSS '16), 1909-1918. http://dx.doi.org/10.1109/HICSS.2016.242
36. Jennifer Shirk, et al. 2012. Public participation in scientific research: A framework for deliberate design. *Ecology and Society* 17, 2, 29-49.
37. Daniel Solove. 2008. *Understanding Privacy.* Harvard University Press.
38. Frederic Stutzman and Woodrow Hartzog. 2012. Boundary regulation in social media. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (CSCW '12), 769–778. http://dx.doi.org/10.1145/2145204.2145320
39. Brian Sullivan, et al. 2014. The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation* 169, 31-40.
40. Latanya Sweeney. 2002. K-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, 5, 557-570.
41. Jessica Vitak, Katie Shilton, and Zahra Ashktorab. 2016. Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW 2016), 941-953. http://dx.doi.org/10.1145/2818048.2820078
42. Andrea Wiggins. 2013. Free as in puppies: compensating for ICT constraints in citizen science. In *Proceedings of the 2013 Conference on Computer*

Supported Cooperative Work (CSCW '13), 1469-1480. http://dx.doi.org/10.1145/2441776.2441942

43. Andrea Wiggins and Kevin Crowston. 2012. Goals and tasks: Two typologies of citizen science projects. In *Proceedings of the 45th Hawaii International Conference on Systems Science* (HICSS '12), 3426-3433. http:// http://dx.doi.org/10.1109/HICSS.2012.295
44. Andrea Wiggins and Yurong He. 2016. Community-based data validation practices in citizen science. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16), 1548-1559. http://dx.doi.org/10.1145/2818048.2820063