Human Computation (2015) 2:1:19-44 © 2015, Bowser & Wiggins. CC-BY-3.0

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

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

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 di-

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

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 especial- ly participant primacy	<i>Respect for persons</i> and <i>participant primacy</i> suggest that par- ticipants have the right to determine what happens to them through informed consent. <i>Data legibility</i> and <i>longitudinal</i> <i>engagement</i> advocate that volunteers should be continually informed about the uses of their data and their rights as con- tributors.
Benefi- cence	Data legibility	<i>Beneficence</i> suggests that benefits to participation should be maximized, and <i>data legibility</i> suggests that participants may benefit from access to data and analysis opportunities.
Justice	Longitudinal en- gagement	<i>Justice</i> advocates that risks and benefits should be balanced. <i>Longitudinal engagement</i> suggests that volunteers should con- tinually 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 poli-

cies 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 phe

ever. 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 par- ticipants, 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. Some projects collected potentially personal information, such as location, without explicitly mentioning it in their policies.		

Data management

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 specif-

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 few⁸. 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

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

Belmont

<u>Table 3.</u> 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).

tion more broadly. In addition, these practices can support stronger relationships between projects and their volunteers, and between different projects.

5.3.

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

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 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

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

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. *Communica*tions 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.

9. APPENDIX

Project Desc	ription	Participation model	Primary project goals	Status
--------------	---------	------------------------	--------------------------	--------

CS1

!