Public policy plays an influential role in the work we do as HCI researchers, interaction designers, and practitioners. "Public policy," a broad term, includes both government policy and policy within non-governmental organizations. This forum focuses on topics at the intersection of human-computer interaction and public policy. **— Jonathan Lazar, Editor**

# Sharing Data While Protecting Privacy in Citizen Science

**Anne Bowser,** University of Maryland and Woodrow Wilson International Center for Scholars Commons Lab, **Andrea Wiggins,** Cornell Lab of Ornithology and University of New Mexico, **Lea Shanley,** Woodrow Wilson International Center for Scholars Commons Lab, **Jennifer Preece,** University of Maryland, **Sandra Henderson,** Project Budburst

Imagine finding a flower you haven't seen before, or watching the first swallows arrive in your garden each spring. How would you record and share this data with those who can advance science or shape public policy? Every day, millions of "citizen scientists" participate in research to support real-world goals [1]. eBird volunteers contribute natural observations of birds that ultimately inform land conservation policies (http://ebird.org). Players of Foldit, an online game for citizen science, helped to identify a protein crucial to the reproduction of HIV (http://fold.it).

Traditional scientific research has well-established rules and procedures, such as informed consent, to ensure the privacy and security of individual participants. In contrast, citizen science projects frequently are created by community members who come together with shared concerns, such as air pollution or loss of habitat. These projects use data contributed by members of the public. Consequently, citizen science practices and the technologies that support them may be designed without privacy in mind. But in these cases, as with scientific research, protecting participant privacy should still be a key concern.

Consider a hypothetical example: Track-a-Tree (TaT) is a popular online citizen science project that asks people to find a deciduous tree and report its seasonal changes. Individuals register their tree and its location online. Participants submit observations and photos of each tree to the TaT website. TaT has an open access policy, so all data is freely available.

TaT is concerned that personal information about participants may become available on its website. In addition, uploaded photographs may include recognizable people, depicting whom they were with and where they were. TaT could have legal problems if such images (especially of children) are published without proper permissions.

Because each tree's location is made public, TaT staff worry that their database may convey the locations of participants' homes, schools, workplaces, or places of worship. This could enable others with malicious intentions to find participants, and might reveal sensitive information about participants' affiliations. As an educational project, TaT also must ensure that children under 13 are guided by a responsible adult.

While TaT is a hypothetical project, many of these concerns—including safety from harassment and stalking, the confidentiality of personal information, and civil liberties relating to the protection of individual privacy—are very real. Here we define privacy as "the right to manage access to voluntarily submitted personal data." Personal data refers first to personally identifiable information (PII). As defined by the Office of Management and Budget, PII includes information that can distinguish an individual (such as full name) and information that can be linked to an individual (such as medical records or IP address). We also consider personal data information about volunteers that is embedded in their submissions. For example, geotagged observations reveal the location of an observed specimen and also that of the volunteer documenting it.

Examples from two citizen science projects ground our discussion. Project BudBurst engages participants to collect data about the timing of plants' leafing, flowering, and fruiting (www.budburst.org). This data is collected in a standardized manner so that scientists can study the responsiveness of plants to changes in climate. iNaturalist is an online community of nature enthusiasts who submit "casual" and "research grade" observations of plants, animals, or fungi (www.inaturalist.org). Scientists use research grade data from iNaturalist, while volunteers use the platform to manage their own data, explore others' observations, and interact through direct messaging or following one another's online activity.

Through these examples, we explore legal and policy

**Insights**

- → Citizen science projects must protect the privacy of volunteers by informing them about potential threats and implementing safeguards.
- → Privacy laws, policies, and standards exist to guide developers of citizen science applications toward best practices.

considerations protecting participant privacy in citizen science and how these considerations relate to technology design. In general, projects can develop policies that incorporate laws and ethical standards into organizational practices. These policies support privacy by allowing users to make informed decisions. In other cases, solutions are designed into technologies, with automated processes protecting volunteers, to greater or lesser extents, regardless of whether users realize it.

## LEGAL ISSUES

Laws are enforceable rules that govern behavior. In the U.S. there are only a handful of federal laws that protect the privacy of individuals, including those engaged in citizen science, summarized in Table 1. (Please note that our discussion prioritizes U.S. law and policy for the sake of focus and brevity.)

| LAW | DESCRIPTION | APPLIES TO |
|---|---|---|
| The Children's Online Privacy Protection Act (COPPA) | The online collection of personal information from children under 13 | Many projects, especially those with an educational component |
| Health Insurance Portability and Accountability Act (HIPAA) | Protects the privacy of medical records | Projects in health |
| Privacy Act | Restricts federal agencies from collecting, using, or distributing PII | Projects funded by federal agencies |
| The Freedom of Information Act (FOIA) | Guarantees citizens access to records maintained by the federal government; protects PII from public record | Projects funded by federal agencies |
| E.U.'s Directive on Privacy and Electronic Communications | Websites must clearly state data collection practices, including cookies | Projects doing business in the E.U. |

→ Table 1. U.S. federal privacy laws.

Federal citizen science projects can choose not to collect PII or create a database that cannot be searched by individuals' information. If they do collect PII, they must comply with the Privacy Act by providing information about users' rights and a Privacy Act Statement. The Privacy Act and FOIA dictate that datasets from projects funded by federal agencies must be cleaned of personal information before they can be released to the public.

COPPA limits the collection of personal information from anyone under the age of 13. It applies to citizen science projects across the board, including Project BudBurst and iNaturalist. According to COPPA, the homepage of a website and each page where personal information is collected *must* link to a privacy notice with data policies explaining which data is collected and how; how that data is used; contact information of a person responsible for the data; and parental rights of minors.

Project BudBurst complies with COPPA by providing a link to a Data Policies page describing the act, and by including a statement about compliance in the footer of every Web page. Additionally, volunteers who join Project BudBurst are required to click a box labeled, "I am at least 13 years of age." Volunteers are informed that "[i]f you are under the age of 13, you must have your parent, guardian or teacher register for you." The registration form cannot be submitted if this box is unchecked. Thus, Project BudBurst employs both policy and technological solutions to ensure COPPA compliance.

Because it is impossible to control the content of direct messages, iNaturalist can't ensure that participants under 13 will not share personal information through the website (in violation of COPPA). iNaturalist therefore restricts participation to adults 18 years and older.

## POLICIES AND STANDARDS

Policies are implemented through protocols that guide decisions. Some policies require compliance in specific situations: National Science Foundation grantees must follow a strict set of protocols before receiving funding. *Policy* also refers to guidelines set as standards. While the majority of standards aren't enforced, they reflect a consensus about what is considered acceptable, or benchmarks to strive toward.

In the U.S., the Federal Trade Commission's Fair Information Practice principles offer a set of four privacy standards that include notice, choice, access, and security. A fifth principle, enforcement, often supplements these. The FTC also suggests best practices for supporting mobile privacy, most of which relate directly to one of the principles listed above [2]. Many of these best practices are directed toward "platforms or operating system providers," described as "the interface between users and hundreds of thousands of apps." Others are written for the developers of mobile apps. This implies that ensuring privacy is a shared responsibility.

According to the principle *notice*, "Websites should notify users about their information practices before collecting any personal information." Requiring volunteers to register is a potential barrier to participation, but registration processes create an

opportunity to, among other things, present data policies to potential volunteers. Both Project BudBurst and iNaturalist require volunteers to register, and both projects post links to data policies next to their registration forms (see Figure 1).

Service providers also design policies toward the same end. When users submit data to the Project BudBurst mobile app on an iPhone, they encounter a pop-up that reads: "http://app.budburst.org would like to use your current location" (see Figure 2); Android users receive a similar prompt.

In addition to providing notice of data collection, projects also describe what happens to a participant's data in the event of account deletion. Often, data isn't removed from archival data products in order to support scientific standards of replicability, but rather is not included in subsequent data products.

*Choice* suggests that "users can determine how their personal information is used." Volunteers who submit observations to iNaturalist include coordinates of latitude and longitude (manually, through a website, or automatically, via a mobile device), but have three options for how the location is displayed: open, obscured, or private, with varying specificity. This choice supports privacy, with one key exception: Observations of threatened or endangered species are automatically "private."

*Access* suggests that "users should be able to view and modify information collected about them." In some cases, information "about" a user might refer to PII; other data (such as their locations) might also be embedded in the data they submit. So, in accordance with access, users should be able to view or modify PII collected during registration; this applies to submitted data as well. Project BudBurst provides access to both types of data through the My BudBurst portal, which allows volunteers to view observations, change email preferences, and update membership information from a single page. In other projects, submitted data may be fully editable.

The final principle, *security*, asserts that "websites should take

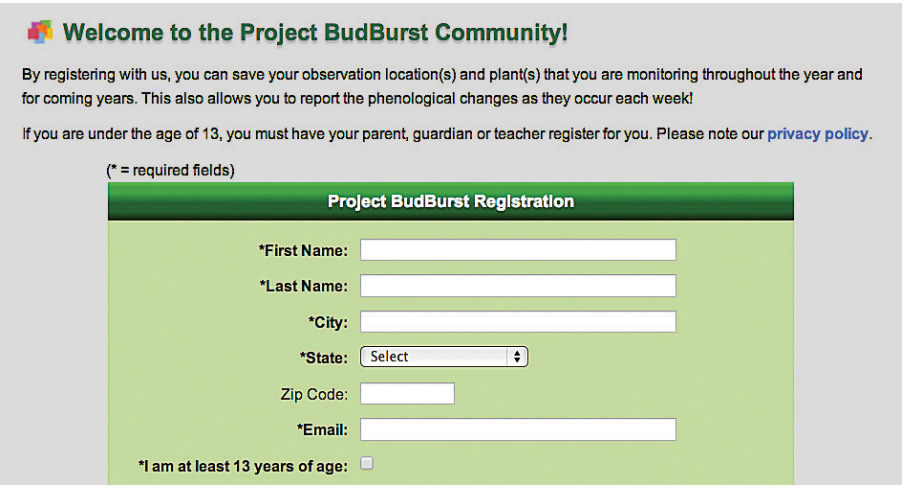


**Figure 1. The Project BudBurst website gives notice by displaying policies during registration.**

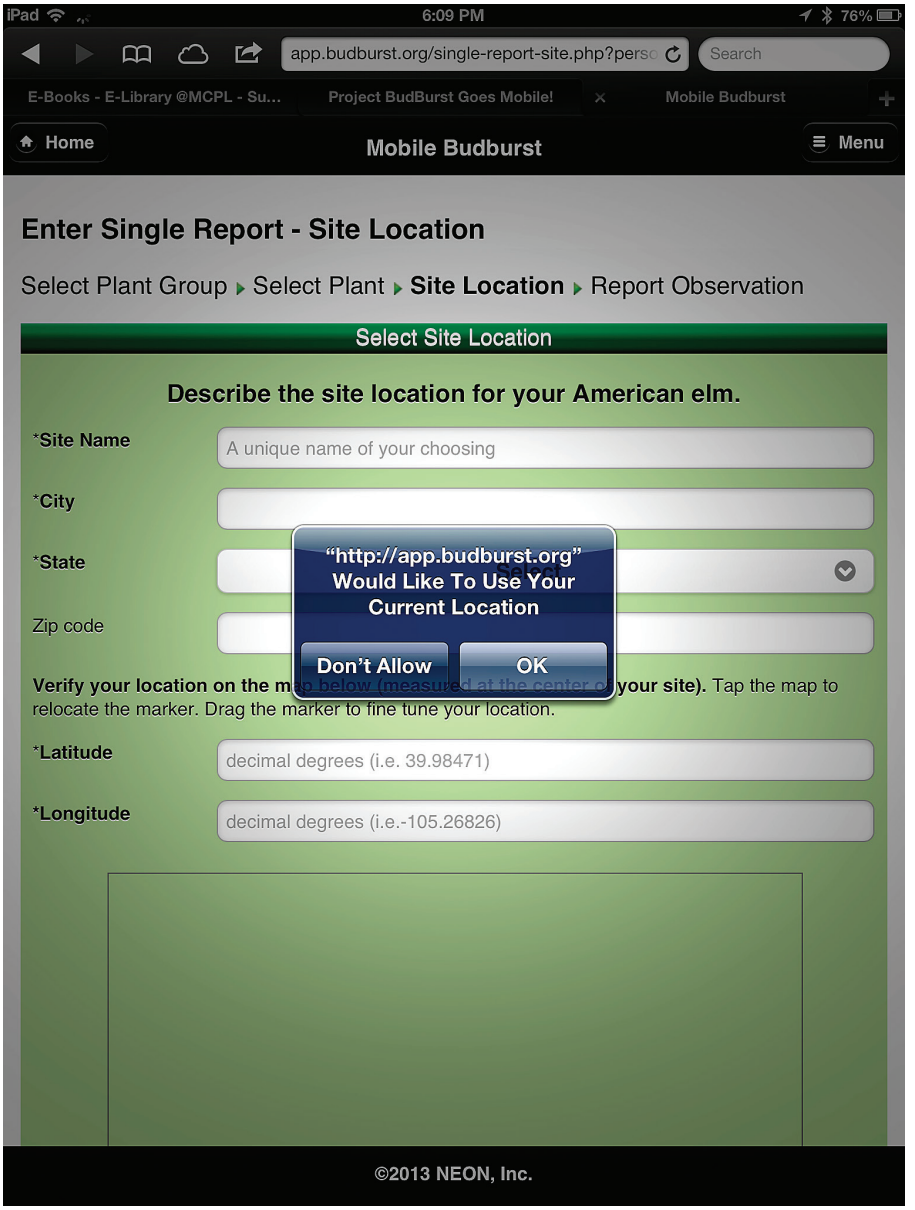


**Figure 2. The iOS platform gives pop-up notices during data submission (a technological solution).**

steps to protect information about their users." Many citizen science projects collect location information that is particularly problematic for privacy and security; recording and displaying locations where data was submitted reveals the position and possibly the movements of the contributor. Additionally, many projects provide open access to project data, including fields such as username, latitude and longitude, and date or time. Some projects use technology to ensure security; one example of this solution is the "fuzzing" of locations in iNaturalist. Policies also can safeguard security, especially through notice. Project BudBurst reminds volunteers: "Please remember that any information you may disclose in any public areas of our site or on the internet generally becomes public information. You should exercise caution when deciding to disclose personal information in these public areas."

Memorandums submitted by the Office of Science and Technology Policy (OSTP) are additional sources of U.S. policy. For example, one memorandum titled "Expanding Public Access to the Results of Federally Funded Research" asks agencies to maximize access to published data while "protecting confidentiality and personal privacy." Standards outside the U.S. include principles submitted by the Organization for Economic Co-operation and Development (OECD) [4]. E.U. standards are generally stricter than U.S. standards.

## IMPLICATIONS FOR CITIZEN SCIENCE

Managers of citizen science projects and other crowdsourcing initiatives can design technologies and policies to support privacy in a number of ways. In addition to the guidelines proposed by the FTC, we propose the following best practices:

- Determine which data points you can and cannot compromise on in terms of precision, public visibility, and data sharing; clearly state these decisions, and implement the supporting technologies (fuzzing locations, anonymizing identities, etc.).
- Give ample notice of privacy choices. Explain the circumstances under which normal participation could be a risk to personal privacy. Inform volunteers who will review their data for quality control.
- Give volunteers the option to hide certain data points and locations from public view, or have data publicly visible but attributed anonymously.
- Allow volunteers to delete and modify their data—both traditional PII and submitted data that may contain information "about" the volunteer.
- Require only minimum personal data about volunteers. Demonstrate the value of the data you collect, and explain who will be able to see it. Multilevel access control that considers different stakeholders' roles and needs may be appropriate.

When we think about privacy, we tend to consider financial or identity theft and putting children at risk; similarly, we natively think of contexts like social media, banking, and online shopping. Although science applications might not spring to mind in relation to privacy, citizen science participants frequently reveal personal details, often unwittingly. Successfully including the public in science requires careful planning, attention to relevant laws and regulations, and adopting appropriate policies and ethical best practices. As leaders, developers, and managers of these projects, we are obligated to protect the privacy of volunteers by informing them about potential threats and implementing safeguards. The thought and skill we put into maintaining privacy will support the success of citizen science.

## ACKNOWLEDGMENTS

We thank Scott Loarie and Carol Boston. This work is supported by NSF Grants OCI-083094, SES-0968546, and by the Commons Lab of the Woodrow Wilson International Center for Scholars through a grant from the Alfred P. Sloan Foundation. Project BudBurst is funded by NSF as part of the National Ecological Observatory Network (NEON).

### ENDNOTES

1. Cohn, J.P. Citizen science: Can volunteers do real research? *BioScience 58*, 3 (2008), 192–107.
2. See http://www.ftc.gov/os/2013/02/130201mobileprivacyreport.pdf
3. See http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
4. See http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

**Anne Bowser** is a Ph.D. student at the University of Maryland's Information School, where her research focuses on gamification in citizen science. She also works as a graduate research assistant at the Woodrow Wilson Center's Commons Lab, and as a student researcher to the ACM-SIGCHI Project on Human-Computer Interaction Education.
→ anne.bowser@wilsoncenter.org

**Andrea Wiggins** (andreawiggins.com) is a postdoc with DataONE at the University of New Mexico, co-appointed with the Cornell Lab of Ornithology. She has a Ph.D. in information science and technology, an M.S.I. in information, and a B.A. in mathematics. Her research focuses on citizen science and open participation systems.
→ andrea.wiggins@cornell.edu

**Lea Shanley** directs the Commons Lab within the Science and Technology Innovation Program (STIP) at the Woodrow Wilson Center. The Commons Lab seeks to advance research and independent policy analysis on emerging technologies that facilitate collaborative, science-based, and citizen-driven decision-making, with an emphasis on their social, legal, and ethical implications.
→ lea.shanley@wilsoncenter.org

**Jennifer Preece** is professor and dean at the University of Maryland's Information School. She authored or coauthored three high-impact books: *Human-Computer Interaction* (1994), *Online Communities: Designing Usability, Supporting Sociability* (2000), and *Interaction Design: Beyond Human-Computer Interaction* (2002, 2007, 2011). She researches the motives of citizens to contribute citizen-science data.
→ preece@umd.edu

**Sandra Henderson** is director of NEON Citizen Science in Boulder, CO. She is a co-founder and director for Project BudBurst (www.budburst.org) and is active in climate change education. She was recently honored as a White House Champion of Change for her work in the field of citizen science.
→ shenderson@neoninc.org

DOI: 10.1145/2540032 