Data Policies for Public Participation in Scientific Research: A Primer

DataONE Public Participation in Scientific Research Working Group
August, 2013

Anne Bowser, Andrea Wiggins, and Robert D. Stevenson

Abstract

This guide introduces data policies in the context of Public Participation in Scientific Research, or citizen science. Four types of data policies are examined: user agreements, terms of use, legal policies, and privacy policies. Each is briefly explained with examples from existing citizen science projects and best practices for implementation. Practical instructions that describe how to craft data policies for citizen science projects and information on how data policies fit within the data management lifecycle are included. This guide was designed to inform project managers about data policies for their citizen science programs. Citizen science volunteers will also benefit from this guide by learning about the rights and responsibilities that their contributions entail.

Introduction

Public participation in scientific research (PPSR), commonly known as citizen science, is a type of collaboration where volunteers work with professional scientists to study real world problems (Cohn, 2008). As such, citizen science combines research with opportunities for education and the possibility to inform different policies (Roy et al., 2012). While sharing a general model of public participation in scientific research, citizen science projects differ from one another in many important ways. For example, projects differ in the level that they involve volunteers in the research process (Bonney et al. 2009). Some projects ask volunteers only to contribute scientific data such as observations. Other projects encourage volunteers to play a more central role in choosing research questions, refining a research design, analyzing data, and/or disseminating interpreted results.

Projects also differ in their organizational structure, and support different types of participation (Wiggins & Crowston, 2008). Some projects enable volunteers to influence their local communities or address civic concerns through the use of scientific data. Other projects ask participants to collect data in order to answer specific research questions. Projects also use different tools for data collection, such as paper data forms or online checklists. Finally, projects focus on different data types and sciences. Some projects study life on earth: these range from annual bird counts, to plant phenology monitoring, to investigations into the shapes of microscopic cells. Other projects study environmental forces such as weather. Still others analyze the structure of galaxies or proteins.

All of these variables—level of participation, organization, data collection, type of data—determine how projects are planned and implemented, and how data are collected, preserved, and analyzed. Because of variation among projects, no single data policy can meet the requirements of every citizen science project. Rather, this guide is designed to meet three primary goals:

1. Introduce data policies by identifying the types of policies most relevant to citizen science projects

2. Illustrate how a project's unique properties and constraints determine which policies are adapted by highlighting the policies of different citizen science projects

3. Empower projects to develop strong, appropriate data policies through the achievment of the first two objectives

About this guide

This guide is structured around the metaphor of a data policy puzzle (Figure 1). Jigsaw puzzles form a picture from interlocking pieces. Similarly data policies are complex documents that contain interlocking types of information. The puzzle piece metaphor was selected because data policies may be easier to understand if similar types of policies are grouped together as adjacent pieces.
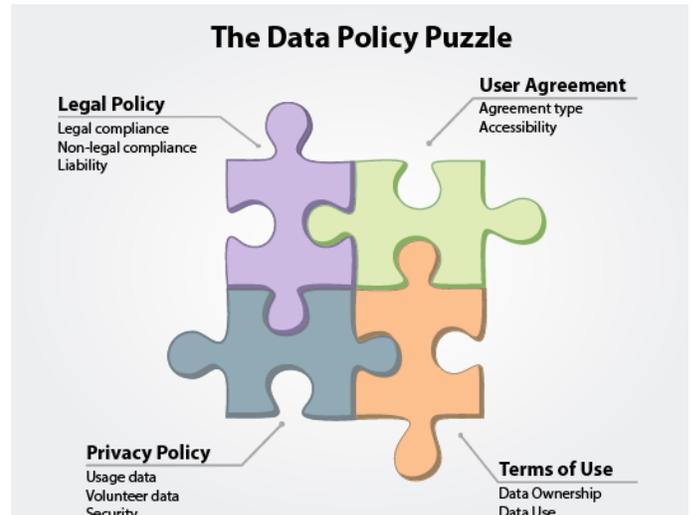

Figure 1. Visualizing the different pieces of the Data Policy Puzzle

It may also be the case that practitioners skimming this guide realize early on that their projects do an excellent job addressing some types of data policies but would benefit from re-evaluating their approach to other types of data policies. With this in mind, the guide is designed to support different types of use. Practitioners who already have strong data policies can use the puzzle piece framework to identify potential areas of weakness (such as legal agreements), and then read the sections for each of these areas in detail.

Those seeking to evaluate existing policies could also skim through the guide for the projects highlighted in pink, and read the information pertaining to projects that are similar to their own. Projects that are building their data policies are advised to read the entire guide if possible and, if not, to begin with the best practices included in each section and then read the sections that seem to be most important to dealing with a unique set of constraints are most confusing.

Different types of data policies are often placed on different web pages. For example, iNaturalist links to a "Terms of Service" page and a separate "Privacy Policy" page in the footer of their homepage. As Nature's Notebook demonstrates, a single page can effectively contain multiple types of data policies. However, it's important to note that no piece operates in isolation. As with putting together a physical puzzle, there is no specific order required in constructing a data policy. Any of the pieces can be put into place at any given time.

This guide begins by explaining the data policy puzzle, situating data policies in the context of the data life cycle, and defining key terms. Next, each puzzle piece is presented in its own section. These sections include general guidelines, examples from ongoing projects, and best practices.

After the conclusion of these guidelines, references are listed for learning more about the different aspects of data policies. Specific citizen science projects that are referenced as exemplars are also listed.

Disclaimer

This primer represents the authors' understanding of data policy in citizen science. It does not represent the beliefs of any organizations affiliated with the authors, including DataONE and the National Science Foundation. Information included in this guide, including information on legal policy, is not legal advice. This guide cannot be substituted for the advice of a licensed legal professional.

The Data Policy Puzzle

Data policies can be considered collections of guidelines that determine how one party, such as a citizen science project, can interact with another party, such as a website user or a citizen science volunteer. They include four main components: user agreements, legal policies, privacy policies, and terms of use.

1. *User Agreements* are contracts between a project and a user that outline the acceptable behavior of both parties. User agreements cover both online and offline behavior.

Data Policies: A Working Definition

Data Policies can be considered a collection of guidelines that determine how one party, such as a citizen science project, can interact with another party, such as a website user or a citizen science volunteer. They are usually present in one or more formalized policy documents. These include:

1. **User agreements**
2. **Terms of use**
3. **Legal policies**
4. **Privacy policies**

Key Terms Defined

Project: a citizen science organization, program, specific campaign, or other entity that solicits participation from volunteers in scientific research activities

User: any person who accesses the website, scientific data, or other materials presented or generated by a project

Volunteer: any person who submits observational data or otherwise contributes to a project, or expresses intent to participate through registration

---

- A clickwrap agreement is a type of agreement that requires a user to explicitly agree to abide by a set of guidelines, such as clicking a box that says "I agree".

- A browsewrap agreement is a type of agreement that does not require a user to explicitly agree to abide by a set of guidelines; rather, a user accepts these guidelines simply by using the website.

2. *Terms of Use* dictate how a website and its contents can be used. They include information on data ownership, data access, data reuse, citation, and other forms of attribution.

3. *Legal Policies* describe how a project adheres to relevant national and local laws, and may also include ethical guidelines.

4. *Privacy Policies* describe how a project gathers, discloses, and manages user and volunteer data, including the use of data gathered by cookies and web logs, personal data gathered during registration, and data contributed by volunteers.

Some topics included in this primer, such as copyright, could easily be included in multiple sections of the primer (in this case, "Legal policies" or "Terms of use"). And in fact, some policies—such as The Children's Online Protection Act (COPPA)—are sufficiently important to be discussed in multiple areas.

User Agreement

A user agreement is a legally enforceable contract between a project and a volunteer who provides observations or other data. When a volunteer accepts a project's user agreement, both the project and the volunteer agree to abide by the project's legal policies, privacy policies, and terms of use.
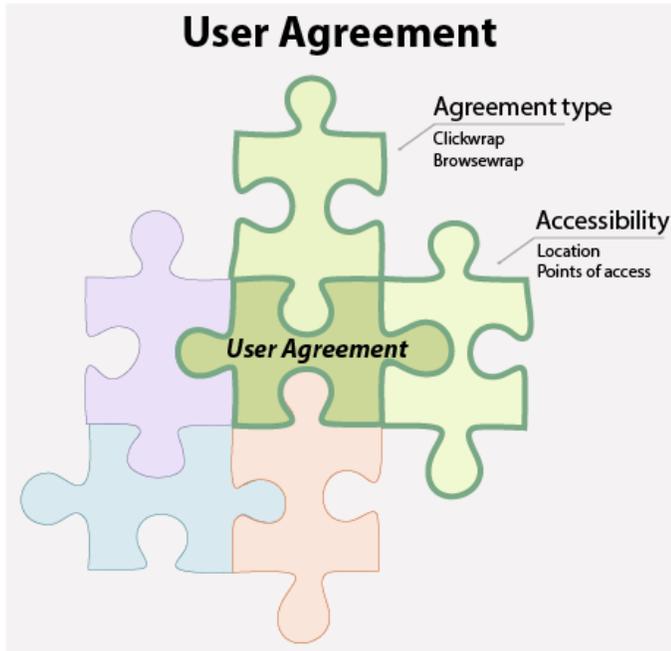

Figure 2. Pieces of a user agreement

Figure 3. The Encyclopedia of Life's Clickwrap Agreement

As illustrated by Figure 2, agreement type and accessibility are two key factors that shape a user agreement.

Agreement Type

There are two main types of user agreements: clickwrap, and browsewrap. Clickwrap agreements are explicit agreements: a user actively consents to the terms outlined in website policies. Some websites embed their clickwrap agreements in a form, and require a user to scroll through the entire agreement before they are able to click an "I agree" button. This is a common practice for websites operating in high risk conditions, such as software providers that install new programs on a user's hard drive. For citizen science projects, it is much more common to require users to click a check box with a label such as "I agree to the terms of use." Often, this is linked to a web page with a project's terms of use or a privacy policy for that website. The Encyclopedia of Life (EOL) structures its user agreement with a clickable link (Figure 3).

In contrast to clickwrap agreements, browsewrap agreements are implict. Projects that use browsewrap agreements still put their policies on a website for volunteers to read; however, users do not have to click a box or otherwise indicate agreement in order for it to take effect. For example Mountain Watch (governed by the Appalachian Mountain Club, or AMC) describes its browsewrap agreement as: "By using this site, you consent to the collection and use of this information by the AMC." A sample browsewrap agreement from the creative commons is displayed in Figure 4.

Generally, courts are likely to agree that clickwrap agreements provide users with sufficient ability to review their data policies. One early (1998) ruling, *Hotmail Corporation vs. Van Money Pie., Inc.*, found that the opportunity to review an agreement coupled with the condition of clicking "I agree," was sufficient to indicate legal consent (Samson, 2008a). Rulings related to browsewrap agreements are mixed. For example, a 2000 ruling on a complaint leveraged by Ticketmaster.com against Tickets.com found that the simple act of linking to terms and conditions in the footer of a homepage was insufficient to constitute a binding agreement, but this ruling was later reversed (Samson, 2008b).

Figure 4. Example of a Browsewrap agreement for the use of the Creative Commons website

Accessibility

The placement and type of a user agreement on a project's website, in terms of the location of a user agreement on a project's website, are both important factors that can affect legal enforcement. A "genuine" clickwrap agreement can be considered one where the entire terms and conditions are located next to a click-button or check-box prompting a user agreement (Bayle, 2009). Rulings such as *Sprecht v. Netscape* generally find these agreements legally binding. In contrast, courts may be less likely to find agreements that simply link to the full text of a user agreement to be valid.

The reason for this discrepency is the idea that a user needs to be provided with "reasonable notice and opportunity to review" an agreement before providing their informed consent. Notably, providing "reasonable notice" does *not* suggest that a user has actually read or reviewed the documents that were shown. In fact, critics of reasonable notice complain that most sensible readers will not spend time reading a lengthy list of terms and conditions (Looijen, 2010).

Best Practices: User Agreement

- **Present** volunteers with a user agreement and associated policies prior to registration
- **Place** the user agreement on a single page, with no scroll box
- **Provide** an opportunity to print and/or save the user agreement
- **Offer** both the options to accept and decline a user agreement
- **Ensure** that the user agreement is posted permanently on the site

(adapted from Bayle, 2009)

Terms of Use

Terms of Use dictate how a website and its contents can be used. These include ownership or copyright of individual pieces of data, ownership or copyright of a data set, policies for data access, and policies for data use (Figure 5). Terms of Use (TOU) are often hosted under titles such as Terms and Conditions, Terms of Service (TOS), Data Use, or Use Policy.

Data ownership

Data ownership is claimed as a type of intellectual property right. Intellectual property is a legal concept that deals with intangible creations of the mind such as creative or intellectual works. Types of intellectual property rights include trademarks, patents, and copyright. Essentially, these rights state that the person who created a piece of intellectual property owns that property, and is solely entitled to the benefits associated with ownership and use. As such, intellectual property rights were designed to incentivize innovation by allowing for adequate reward. Copyright, which protects the form or expression of an idea, is the type of intellectual property most frequently
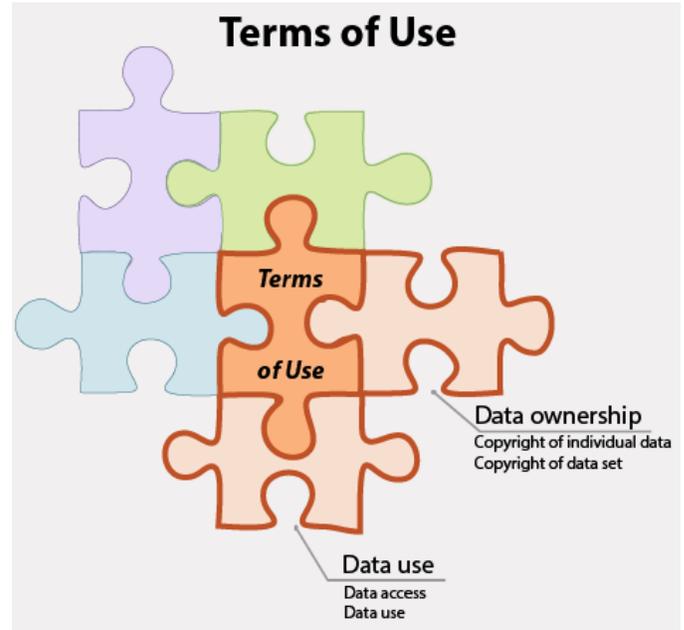
Figure 5. Pieces of Terms of Use

discussed in citizen science. Specifically, two types of copyright are important to examine: the copyright of scientific data, and the copyright of databases.

When scientific data are collected, the scientist or investigator initially owns the copyrights to that data. Historically, a scientist would waive their copyright of both data and article to a publisher or other distributor during the publication process. This transfer of rights is considered justified by the benefits that a scientist gains from the publication and subsequent dissemination of their work (Elliot, 2005). However, these practices are changing. For example, *Nature* mandates that "a condition of publication in a *Nature* journal is that authors are required to make materials, data and associated protocols promptly available to others without undue qualifications." This condition is maintained so that others can replicate and build upon scientific work.

Databases are described as collections of data that are organized in a systematic way (Elliot, 2005). In European law, the effort involved in compiling data into a database is sufficient to grant copyright to the database owner. This was also the case in the United States prior to the 1991 ruling *Feist Publications v. Rural Telephone Service Co.(24)*. *Feist* found that the "sweat of the brow" of a database compiler was insufficient to merit copyright; instead, the court confirmed that "independent creation plus a modicum of creativity" was required (U.S. Copyright Office, 1997).

The modicum of creativity can be small: for example, a compilation of local restaurants that only includes restaurants with high reviews would probably be sufficient (Elliot, 2005).

Citizen science projects that claim copyright to all of the data they hold can do so under copyrights involving "independent creation plus a modicum of creativity."

Calflora is one project that claims copyright of all data:

> "Customer acknowledges that the data are proprietary to Calflora and Calflora's licensors, that the data shall remain the exclusive property of Cal flora or its licensors, and that Customer has no rights therein other than as set forth herein. The data are copyrighted by Calflora or its licensors, and all rights are reserved."

Other projects adopt a more nuanced policy. Zooniverse, for example, leaves copyright of individual *data* to the author or investigator who creates it: "You retain ownership of any contribution you make to the Zooniverse...you may use, distribute or modify your individual contribution in any way you like." However, Zooniverse subsequently retains copyright of the full data set contained in its *database* by adding, "you do not possess ownership of the dataset itself."

Lastly, some projects—especially those that curate the work of others—use their data policy to address unintentional violations of the copyright of other parties. For example, iNaturalist includes the following clause:

> "As iNaturalist asks others to respect its intellectual property rights, it respects the intellectual property rights of others. If you believe that material located on or linked to by iNaturalist.org violates your copyright, you are encouraged to notify iNaturalist. iNaturalist will respond to all such notices...by removing the infringing material."

Data use

If a project claims copyright of some or part of the data that it gathers, then the project has the right and the responsibility to determine how others can use that data. This involves detailing the conditions for accessing and using the data. Data access can be viewed on a spectrum from very limited to very permissive. Many citizen science projects allow limited access to the data that they gather,

---

Best Practices: Terms of Use

- **Determine** who holds copyright of data submitted by volunteers
- **Determine** who holds copyright of the aggregate data set
- **Post** the copyright permissions for individual data and the data set in your user agreement and on your policies page
- **Determine** how you want your data to be used
- **Post** explicit permissions for access and reuse including, if applicable, a citation to the data set or a related paper, community attribution, and requirements for co-authorship
- **Include** attribution and license information in downloaded data files

---

or provide different levels of access based on a user's relationship with the project.

In many cases, projects present data in aggregate form. Mountain Watch, for example, shares data primarily through a newsletter emailed to participants. ReClam the Bay hosts annual reports containing interpreted data and journals of project activities on their website, but does not offer participants access to original data. In contrast, projects such as Project Budburst host both spreadsheets of raw data and PDFs of aggregate summary reports on their project website. Still other projects, such as eBird, make their raw data accessible on their own website and in other open access data repositories. It is important to note that projects adopt different policies for a user's access to the data he/she submits, and for access to an entire data set.

The decision to grant access to raw or aggregate data sets is one that each project must make individually. With that said, programs that wish to share their data in various forms must have both the *intent* to share data and the *ability* to. Compiling and interpreting data for annual reports and newsletters takes time, as does adding sufficient metadata to make raw data sets valuable. Technological limitations and organizational constraints can also inhibit data sharing (Sayogo & Pardo, 2012).

As with policies regulating access to data, policies regulating the use of data fall along a spectrum from very limited to very permissive. Some projects do not allow volunteers to use their data except in very specific conditions. Explaining such a policy, The Great Sunflower Projects writes:

> "Because of the work that goes in to collecting and verifying the data, publication of descriptive or interpretive results derived from the data is the privilege and responsibility of the investigators who collect the data...any researcher making substantial use of these data must communicate with the Great Sunflower Project prior to publication. In most cases, these researchers should anticipate that someone from the Great Sunflower Project would be co-authors of published results."

This final clause, suggesting potential for co-authorship, is shared by many other projects such as eBird.

Duke and Porter offer a good starting point for understanding co-authorship and data reuse (2013). They suggest four criteria for determining whether an offer of co-authorship is reasonable:

> 1. Are the data integral to analysis?
> 2. Are the data novel or unique?
> 3. Is the data provider willing to share authorship?
> 4. Is the provider able to participate?

If all four criteria are satisfied, data providers should be invited as co-authors. If not, it may be more appropriate to cite the data used (assuming that the project permits data use without co-authorship).

Other projects provide unrestricted access, allowing anyone to aquire, analyze, interpret, or redistribute their data. For example, Project Budburst data "is freely available for anyone to download and use." Projects that allow for such data use typically contain conditions for citation or attribution; Project Budburst notes,

> "the data is provided by thousands of observers from across the country. If you use data submitted by Project BudBurst observers for analysis, reports, or presentations, we ask that you link to our Community Attribution page to recognize the efforts of our dedicated volunteers."

Other projects only share raw data after the data are analyzed and interpreted in a scholarly publication (Duke & Porter, 2013).

Creative Commons De-mystified

Creative Commons (CC) was developed as an alternative to traditional forms of copyright. There are two main types of CC licenses: those that reserve some rights, and those that do not reserve any. Six licenses reserve some rights. Of these, the licenses used most frequently by citizen science projects are:

- *Attribution:* others may share and modify your work, even for commercial purposes, but must give you credit

- *Attribution-NonCommercial:* others may modify your work, except for commercial purposes, but must give you credit

- *Attribution-NonCommercial-ShareAlike:* others may modify your work, except for commercial purposes, but must give you credit and must license their work under identical terms

Occasionally, the creator of a piece of copyrighted intellectual property decides that they want to relinquish their rights to the public domain. However, once copyright is granted it can be difficult to legally revoke.

The Creative Commons also offers a *No Rights Reserved (CC0)* license that allows people to waive all interests in their work, placing it "as completely as possible" in the public domain.

Many projects that allow re-use have adopted Creative Commons licenses as a method of clearly explaining their policies. In some cases, a single license applies to all of a project's data; in other cases, different licenses apply to different types of data. For example, Great Worldwide Star Count licenses their data under a general Attribution-ShareAlike license, but licenses their classroom activities and teaching resources under the more specific Attribution-NonCommercial-ShareAlike license. This means that their data can be used for commercial purposes, but their teaching tools cannot.

Legal Policies

Legal policies outline the local laws, state laws, national laws, and other guidelines that govern the interactions between citizen science projects and their volunteers. This document primarily discusses federal regulations of the United States.
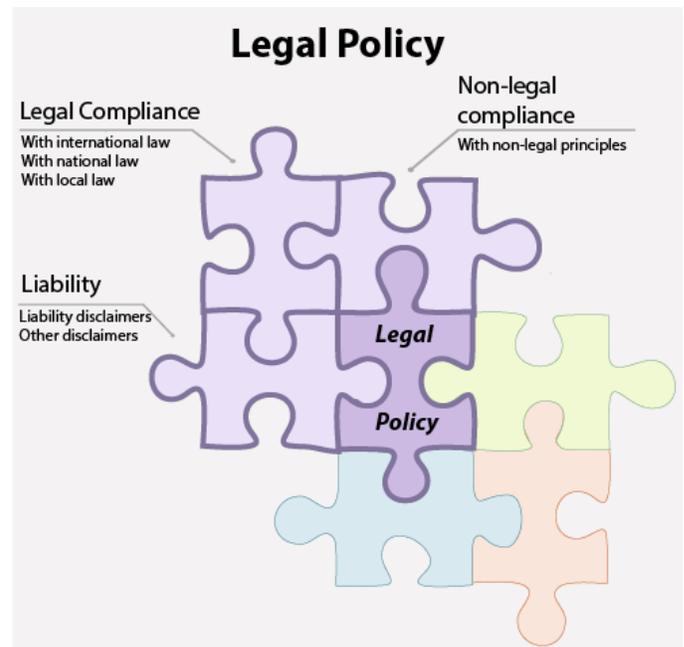

Figure 6. Pieces of Legal Policy

Some legal policies exist to protect the rights of volunteers who contribute their data. These include policies relating to COPPA, which dictate the conditions where data can be collected from children under 13. Other policies, such as those that limit the liability of different projects, are designed to protect the projects themselves.

As illustrated by Figure 6, there are three main types of legal policies relevant to citizen science. One important type of legal policy outlines compliance with local, state, and national law. A related type of policy is non-legal policy that addresses compliance with standards such as The US Department of Health and Human Service's Internal Review Board (IRB) guidelines dictating acceptable conditions for involving human subjects in research. Finally, liability policies explain a project's legal obligations and related disclaimers.

The Children's Online Privacy Protection Act (COPPA), which outlines the conditions for collecting personal data from children under 13, is a crucial legal consideration. Some projects, such as the Great Worldwide Star Count, explicitly list COPPA compliance: "We ensure that our privacy policy and our information practices adhere to the United States Children's Online Privacy Protection Act, and other applicable data privacy laws." Others, projects, such as the Encyclopedia of Life (EOL), simply note that they do not "knowingly collect personally identifiable information from children under 13." COPPA dictates that a privacy policy must describe "the kinds of personal information collected from children...and how the information is collected" (http://www.coppa.org/comply.htm).

Table 1. Legal and other institutional policies relevant to citizen science

Policy	Applies to	Description
Legal regulations in the United States		
Children's Online Privacy Protection Act of 1998 (COPPA)	**Commercial** websites that collect personal data from children under 13	The homepage of a website, and each page where personal information is collected, must link to privacy notice containing: - The types and methods of personal data collection - How personal data are used or shared with other parties - Contact information of a person responsible for data collection - Parental rights pertaining to minor data collection and use
Controlling the assault of non-solicited pornography and marketing (CAN-SPAM)	Entities that use email for marketing purposes	Unsolicited marketing emails may be sent if: - An operatable "unsubscribe" button is included in all emails - The address of a publisher or advertiser is present - Message is sent directly from a sender to a known recipient (i.e., robots may not be used)
Legal regulations outside the United States (example)		
Directive on Privacy and Electronic Communications (E-Privacy Directive)	Any business with users or customers in the EU	Companies in the EU must comply with the E-privacy directive; many companies outside of the EU consider its provisions best practices. Provisions include: - Users must be notified if a website uses cookies - Users must "opt in" to receiving marketing emails
Non-legal regulations		
Institutional Review Board (IRB) human subjects research approval	Research with human subjects	In the United States, the Department of Health and Human Services requires IRB approval for research that deals with human subjects, such as: - Conducting surveys, interviews, or observational research with volunteers - Research involving the collection of personal information
Institutional Animal Care and Use Committee (IACUC) or USDA guidelines for the use of animals in research	Research with vertebrates and other animals, including field studies	The IACUC compiles a list of animal research guidelines and policies at different US colleges and universities. The USDA offers guidelines to conducting research with different types of animals, including birds, fish, mammals, reptiles, and amphibians.

Therefore, projects that explicitly state COPPA compliance should also detail their practices.

Controlling the Assault of Non-Solicited Pornography And Marketing (CAN-SPAM) is a federal regulation that pertains to marketing through mediums such as email. Essentially, CAN-SPAM dictates that while users do not need to "opt in" to a marketing mailing list, they must be able to "opt out" of one. Citizen science projects like What's Invasive demonstrate CAN-SPAM compliance with language like "if you do not wish to be contacted using your email address, please contact us and we will remove it."

The Directive on Privacy and Electronic Communications (E-Privacy Directive) regulates involuntary data collection (such as data collected through the use of cookies) by companies based in European Union member states. Some provisions of the E-Privacy Directive—such as the requirement that users should opt in to marketing email lists, rather than opting out—run contrary to the provisions outlined in CAN-SPAM. Companies based in the United States that might collect data from EU citizens

would be wise to follow these stricter regulations. See Table 1 for a summary of the three guidelines mentioned above, and links to full text documents.

Citizen science projects that are operated by federal agencies may be subject to additional guidelines. For example, The U.S. Privacy Act of 1974 outlines the rights that individuals have regarding personal information collected about them. The Paperwork Reduction Act imposes limitations on who may volunteer for a citizen science project, and outlines the procedures that must be completed prior to the acceptance of new volunteers. Projects run by federal agencies may wish to review the Terms of Use for Nature's Notebook, a federally-funded phenology project with a particularly comprehensive policy that demonstrates compliance with relevant legal policies.

Finally, when writing data policies citizen science projects need to consider relevant state and local law. For example, the Sherman Creek Watershed Monitoring Program reminds volunteers that they must obtain both a fishing

license and a collector's permit to legally collect macroinvertebrates in Pennsylvania. If the Sherman Creek Watershed Monitoring Program did not include this information in their data policy, a volunteer who collected macroinvertebrates without obtaining the required permissions- and fell into legal trouble as a result- could try to hold the program liable.

Non-Legal Compliance

In order to conduct research on human subjects, many projects (including all that are supported directly or indirectly by the Department of Health and Human Services) must gain the approval of an Institutional Review Board (IRB). "Research on human subjects" is broadly defined, and depending on circumstances may include research methods such as surveys, interviews, and gathering observational data. Therefore, projects such as the Appalachian Mountain Club that "may occasionally conduct user surveys to better target our content to our audience" could be subject to IRB guidelines.

To many, the term "research conducted with animal subjects" connotes controlled laboratory studies. However, in reality it can also refer to field studies involving birds, fish, mammals, reptiles, and amphibians. Regulations for conducting research with human and non-human animal subjects are summarized at the bottom of Table 1.

Some projects receive money in the form of grants from federal institutions such as the National Science Foundation (NSF), the National Institute of Health (NIH), or the National Oceanic and Atmospheric Association (NOAA). Such projects may be required to adhere to the legal or ethical requirements of these agencies, which often require more complex plans for data management and sharing. These agencies may also require award recipients to complete IRB training. Links to NSF, NIH, and NOAA guidelines are included in the "Other Resources" section of this primer.

Similarly, projects that are funded by research universities must adhere to university guidelines. Project directors should ensure that their data policies comply with the guidelines stipulated by various sources of funding.

Liability

Liability is a term that refers to legal responsibility. Tort Liability, which occurs when the actions of one group cause injury to another, is the type of liability most closely connected with citizen science and other projects that rely on the efforts of volunteers (Robson, 2012).

Liability becomes important when a project's participation activities harm a person in some way and that person files a legal complaint in response. In the absence of liability disclaimers, projects are vulnerable to being found guilty of harm. However, the simple act of posting a liability disclaimer does not always fully protect a project.

A researcher studying liability disclaimers in sports and recreation suggests that in order for a waiver to be valid,

four conditions must be met (Cotten & Cotten, 1997):

1. A waiver must contain clear and unambiguous language,
2. A waiver must not violate public policy (for example, no waiver will protect a service provider from gross negligence),
3. A user must contractually agree to the condtions of a waiver,
4. A user must be able to contract (for example, they must be of legal age)

Liability complaints are usually heard in state courts. A review of over 900 cases concluded that courts in some states are more likely to enforce liability waivers than courts in other states. Projects might consider using the four conditions described above as a starting point for dealing with liability and then consult a legal professional. In certain conditions other documents (such as informed consent forms) may offer better protection against litigation (Cotten & Cotten, 1997).

In related research, three models of crowdsourcing pose different legal concerns: the unpaid volunteer model, the contest model, and the cognitive-piecework employment model (Frankrone, 2013). The vast majority of citizen science projects follow the unpaid volunteer model, where volunteers share data, advice, or otherwise contribute to an organization in the absence of extrinsic reward. Historically, courts are *least likely* to hold crowdsourcing projects liable for content generated through the unpaid volunteer model.

With that said, certain conditions *increase* the likelihood that a project will be held liable for the content they post. Projects may be more liable for *content that is actively solicited* (Frankrone, 2013). For example, if a citizen science project sends its top users an email commending them for

---

Best Practices: Legal Policy

- **Determine** the relevant legal and non-legal guide lines for your project
- **Indicate** compliance with relevant legal and non-legal guidelines in your data policy
- **Include** a general liability disclaimer along with project-specific disclaimers on your website
- **Consider** purchasing liability insurance that protects board members, employees, and volunteers
- **Generate** general calls for participation instead of soliciting specific types of data
- **Review** as much content as possible before interpreting and distributing aggregated findings
- **Explain** any review processes, including practical limitations
- **Distribute** data documentation that includes legal policies with data files

their contributions and asking for more data, they may be held liable for the content that they receive as a result. Projects that ask for a specific type of data may be similarly liable for what they receive.

Projects can also be held liable for content that is re-distributed. In other words, if a website simply posts the raw data gathered by volunteers, they are acting as a distributor of the data. However, if they aggregate the data gathered by volunteers into an annual report and redistribute it to a broader audience (e.g., by hosting it on their website), or interpret the data in any way, they might be considered a primary publisher or original author, both of which are vulnerable to increased liability (Fitt, 2011). Projects may also be held responsible for content that is reviewed by members of a project team, including comments submitted by users in moderated discussion forums.

Some courts do recognize that holding websites liable for reviewed content would have a negative effect on quality control (Frankrone, 2013). Whether reviewing all content contributed by volunteers is feasible is also an important consideration. Keeping records of decisions related to content review may be useful for this (and other) reasons.

Lastly, citizen science projects may be held liable for harm caused by volunteers. The Volunteer Privacy Protection Act of 1997 (VPA) protects volunteers from liability in the case that they carelessly injure another individual in the course of helping a nonprofit by transferring liability to the nonprofit organization. In other words, an injured party could sue a project but not a volunteer. Liability insurance can be purchased to protect against lawsuits filed against board members, employees, and/or volunteers. However, insured projects may be more likely to be sued than non-insured projects (Robson, 2013)

Citizen science projects address liability in a number of ways. Many projects explicitly state that they are not liable for negative results that arise from using their data. Calflora, for example, "shall not be liable for any loss or damage, including lost profits, loss of use, or incidental, consequential or exemplary damages, caused to any personel resulting from information contained therein." CoCoRaHs adds that it cannot be considered liable for "missing data or misinterpretation." These and similar disclaimers of content liability are very general, and are included in the data policies of most websites within and outside of the citizen science domain.

Many legal companies, findable through a simple search, offer general liability disclaimers that are free to use with attribution. These can provide a good starting point for generating a liability policy, although some projects may wish to include more specific disclaimers. For example, IceWatch USA (which asks volunteers to observe potentially hazardous glaciers) holds volunteers liable for "risks and dangers too numerous to describe" incurred during the process of data collection.

Many organizations offer information on liability as "risk management." The Nonprofit Resource Center offers risk

management tools such as waivers, hold harmless clauses, and a sample volunteer release form.

Privacy Policies

Privacy policies (Figure 7) describe how a project gathers, discloses, and manages user and volunteer data including the use of data gathered by cookies and web logs, personal data gathered during registration, and observational data contributed by a volunteer.
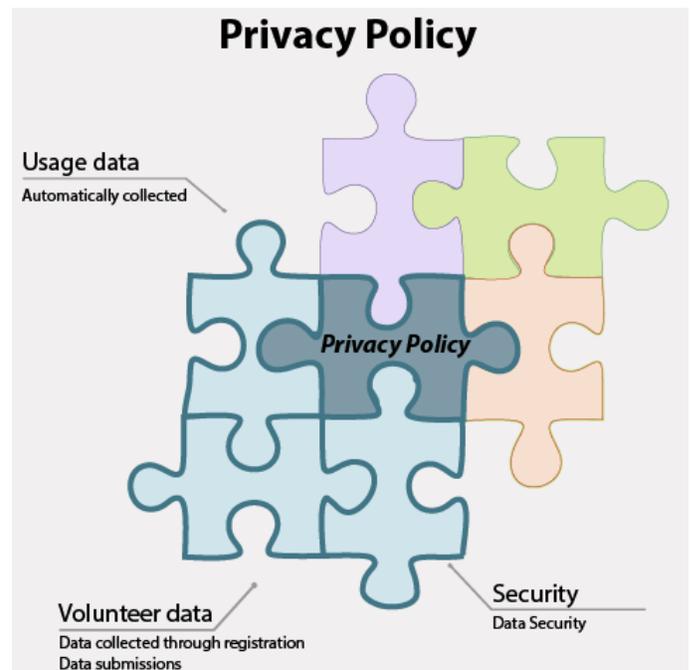

Figure 7. Pieces of Privacy Policy

Privacy policies should be included in a link at the bottom of a home page and all pages where volunteers submit data (Beesley, 2012).

Usage Data

Most websites, by default, gather data about users through cookies and log files. These types of data include service provider, operating system, browser, IP address, date and time of use, and pages visited. Despite the fact that these are standard practices, many citizen science projects include information about cookies and log files on a "Privacy policy" page. As eBird explains,

> "Cookies are pieces of data that we store on your computer to tie your information to the eBird program. Cookies allow you to store your login information, so that you will not be required to log in each time you visit the eBird web site. Internet browsers allow you to determine whether or not to accept cookies (check under the preferences or options in the browser menu)."

Including such information adheres to legal regulations such as those discussed in Table 1, builds user trust, and is a best practice in data policies (Beesley, 2012). For example, the Appalachian Mountain Club (AMC) explains,

> "The AMC creates aggregate reports on user demographics and traffic patterns for advertisers,

sponsors and partners. This allows our advertisers to advertise more effectively, and allows our users to receive advertisements that are pertinent to their needs."

Note that data policies should also explain the rationale for gathering and conditions on sharing usage information.

Volunteer Data

In contrast to usage data, which are gathered automatically from every website visitor, additional data about volunteers is collected when they provide it during a registration process. This data can include, for example, the full name of a volunteer, a username, an email address, a physical mailing address, and demographics (if such data could lead to the location or identification of an individual, it is often referred to as personally identifiable information, or PII). Additional data are collected from volunteers who upload observations or contribute data. These types of data include latitude and longitude coordinates, date and time of observation, observational data, and photographs, which may also embed information about camera or smartphone model, software, and other details.

While privacy policies are not required by law, the Federal Trade Commission (FTC) prohibits deceptive privacy policies, including those that collect data or share data in ways that they do not explicitly mention (Beesley, 2012). As noted in the section on legal policy, legal regulations such as COPPA also dicate the types of data that can be collected from children under 13. Two other types of data merit special consideration: email addresses, and location-based data.

Best Practices: Privacy Policy

- **Ensure** that you comply with COPPA if your users could potentially include children under 13
- **Ensure** that you comply with CAN-SPAM if you collect email addresses from volunteers
- **Include** a link to your privacy policy at the bottom of your home page, and each page that collects data
- **Explain** your use of cookies
- **Detail** your use of volunteers' email addresses
- **Detail** your use of any location-based data that you collect; consider allowing volunteers to designate some locations or observations as "private" or otherwise hidden from public view
- **Include** information on data security

Managing volunteer email addresses

Few things annoy volunteers as much as receiving unwanted emails, whether from projects, from related organizations, or from advertisers. Additionally, CAN-SPAM requires that users are able to opt out of receiving emails at any given time (Table 1). Many citizen science projects include information about opting out of emails both in the footer of each email and on a general data

policy page. For example, The Great Sunflower Project writes, "Other people who are signed up can email you from within the website. If you don't want them to do that, you can go to My account, click the edit tab and then uncheck the personal contact form."

Dealing with sensitive location-based data

Location-based data are extremely sensitive and, understandably, volunteers can be hesitant to share them. At a minimum, privacy policies should detail the measures taken to protect location-based data, including listing the parties that have access to this data. For example, What's Invasive includes the following clause in their data policy:

> "When you use What's Invasive!, location information stored by your cellular phone or entered on the website will be transmitted to a secure database. The data collected will be used to generate information about the location and identity of invasive species...researchers will only access the location data to ensure the proper functioning of What's Invasive! and to test the functionality of our software and hardware. We will not analyze your location data to make interpretations about your behaviors without contacting you for permission first, nor will we share this data with others."

iNaturalist takes an even stricter approach. When volunteers add observations to iNaturalist, they are asked to include the location where the observation was made. For each specific observation, iNaturalist allows users to designate the location as open, obscured, or private:

> "• **Open**: Everyone can see the coordinates unless the taxon is threatened.
>
> •**Obscured**: Public coordinates shown as a random point within 10km of the true coordinates. True coordinates are only visible to you and the curators of projects to which you add the observation.
>
> •**Private**: Coordinates completely hidden from public maps, true coordinates only visible to you and the curators of projects to which you add the observation."

Security

Many websites include general assurances of security and specific practices for ensuring security. Great Worldwide Star Count writes:

> "The security of your personal information is very important. We have put in place appropriate physical, electronic, and managerial procedures to safeguard and help prevent unauthorized access, maintain data security, and use the information we collect online appropriately. When you submit sensitive information, such as your credit card number, we encrypt that information using secure socket layer technology."

Many projects also hold volunteers responsible for taking basic measures to protect their security.

As The Encyclopedia of Life (EOL) states, "Unfortunately, no website can fully eliminate security risks. You are responsible for maintaining the secrecy of your registration and password information."

Summary

This introduction to data policies is intended to benefit the organizers of different citizen science projects as they plan, structure, implement, and evaluate their projects. Because data policies detail the rights and responsibilities of both projects and volunteers, the participants of citizen science projects may also find this information useful.

Data policies are complex, and using the metaphor of a puzzle presents different types of policy in smaller sections that are easier to understand. This guide breaks down data policies into four different types of policies: user agreement, terms of use, legal policies, and privacy policies.

It's important to note that no portion of the data policy puzzle exists in isolation. For example, some researchers who study legal policy argue that clickwrap agreements are necessary to make liability waivers valid (Cotten & Cotten, 1997). And, by accepting a carefully written user agreement, a volunteer agrees to abide by a project's privacy policies, legal policies, and terms of use.

We recommend that projects seek legal council to ensure that their policies include all the information necessary to protect both projects and volunteers. The Foundation Center, an organization designed to support philanthropy, provides an excellent curated list of legal organizaitons that specialize in nonprofit law and offer pro-bono legal assistance. As this primer mainly focuses on legal policies within the United States, seeking council is especially important for projects with an international scope.

Acknowledgments

WWW.DataONE.org   info@DataONE.org

@DataONEorg

facebook.com/DataONEorg

Other Resources
- The Children's Online Privacy Act (COPPA). http://www.coppa.org/
- Controlling the Assault of Non-Solicited Pornography And Marketing (CAN-SPAM): http://www.business.ftc.gov/documents/bus61-can-spam-act-compliance-guide-business
- Creative Commons "Choose a License" tool. http://creativecommons.org/choose
- The Directive on Privacy and Electronic Communications (E-Privacy Directive): http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:NOT
- The Foundation Center's list of organizations specializing in nonprofit law: http://foundationcenter.org/getstarted/faqs/html/probono.html
- Institutional Animal Care and Use Committee (IACUC): http://www.iacuc.org/aboutus.htm
- National Institute of Health (NIH) Data Sharing Policy and Implementation Guidelines: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
- National Oceanic and Atmospheric Administration (NOAA) Data Management Planning Procedural Directive: https://www.nosc.noaa.gov/EDMC/PD.DMP.php
- National Science Foundation (NSF) Data Management Plan Requirements: http://www.nsf.gov/eng/general/dmp.jsp
- Nature Policies: Availability of data and materials: http://www.nature.com/authors/policies/availability.html
- Nonprofit Resource Center Risk Management Tools: http://www.nprcenter.org/resource/waivers-and-hold-harmless-clauses
- Privacy Policy Generators: A list curated by The Future of PrivacyForum http://www.applicationprivacy.org/do-tools/privacy-policy-generator/
- Project Open Data: http://project-open-data.github.io/
- US Department of Agriculture guidelines for research with animals: http://awic.nal.usda.gov/research-animals/wildlife-field-studies
- US Department of Health and Human Services (HHS) IRB Guidebook: http://www.hhs.gov/ohrp/archive/irb/irb_guidebook.htm
- US Privacy Act of 1974: http://www.archives.gov/about/laws/privacy-act-1974.html
- US Volunteer Protection Act of 1997 (VPA): http://www.mke-skywarn.org/vpa.htm

Projects Referenced

Calflora: http://www.calflora.org

CoCoRaHas: http://www.cocorahs.org

eBird: www.ebird.og

Encyclopedia of Life: www.eol.org

The Great Sunflower Project: http://www.greatsunflower.org

The Great Worldwide Star Count: http://www.windows2universe.org/citizen_science/starcount

iNaturalist: http://www.inaturalist.org

Mountain Watch: http://www.outdoors.org/conservation/mountainwatch

Nature's Notebook: https://www.usanpn.org/natures_notebook

Project BudBurst: http://www.budburst.org

ReClam the Bay: http://reclamthebay.org

What's Invasive: http://whatsinvasive.com

Zooniverse: http://www.zooniverse.org

Scholarly References

Bayle, E. (2009). The click that binds: Ways users "agree" to online terms and services. Accessed 1 June, 2013 from https://www.eff.org/fa/wp/clicks-bind-ways-users-agree-online-terms-service

Beesley, C. (2012). 7 Considerations for crafting an online privacy policy. *U.S. Small Business Administration*. Accessed 21 June, 2013 from http://www.sba.gov/community/blogs/community-blogs/business-law-advisor/7-considerations-crafting-online-privacy-policy

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., & Wilderman, C. C. (2009). Public participation in scientific research: Defining the field and assessing its potential for informal science education. A CAISE Inquiry Group Report. Washington, DC: Center for Advancement of Informal Science Education (CAISE).

Cohn, J. (2008). Citizen science: can volunteers do real research? *BioScience*, *58* (3), 192-207.

Cotten, D. & Cotten, M. (1997). Legal aspects of waivers in sport, recreation & fitness activities. Canton, OH: PRC Publishing.

Duke, C. & Porter, J. (2013). The ethics of data sharing and reuse in biology. *BioScience, 63* (6), 483-489.

Fitt, V. (2011). Crowdsourcing the news: News organizations liability for iReporters. *William Mitchell Law Review, 37* (4), 1040-1066.

Elliot, R. (2005). Who owns scientific data? The impact of intellectual property rights on the scientific publication chain. *Learned Publishing, 18*, 91-94.

Frankrone, E. (2013). Free agents: Should crowdsourcing lead to agency liability for firms? *Vanderbilt Journal of Entertainment and Technology Law, 15* (4), 883-912.

Looijen, D. (2010). Time for a change: The schema of contract in the digital era. *Journal of Telecommunications and High Technology Law, 8* (2), 547-570.

Robson, E.S., Shanley, L., & Lovell, A. (2012). *Responding to liability: Evaluating and reducing tort liability for digital volunteers*. Washington, DC: Woodrow Wilson International Center for Scholars.

Roy, H., Pocock, M., Preston, C., Roy, D. & Savage, J. (2012). Understanding citizen science and environmental monitoring: Final report on behalf of UK Environmental Observation Framework. Wallingford, NERC/Centre for Ecology & Hydrology.

Samson, M. (2008a). Click-wrap agreements- Internet library of law and court decisions. Internet Library of Law and Court Decisions. Accessed 25 June, 2013 from http://www.internetlibrary.com/topics/click_wrap.cfm.

Samson, M. (2008b). Browse-wrap agreements- Internet library of law and court decisions. Internet Library of Law and Court Decsions. Accessed 25 June, 2013 from http://www.internetlibrary.com/topics/browse-wrap_agree.cfm.

Sayogo, D., & Pardo, T. (2012). Exploring the motive for data publication in open data initiatives: LInking intent to action. In *Proc. of 45th Hawaii International Conference on Systems Sciences*, 2623-2631.

US Copyright Office (1997). Report on legal protection for databases. Accessed 24 June, 2013 from http://www.copyright.gov/reports/dbase.html.

Wiggins, A. & Crowston, C. (2008). From conservation to crowdsourcing: A typology of citizen science. In *Proceedings of HICSS 2004*, IEEE, 1-10.

Wiggins, A., Bonney, R., Graham, E., Henderson, S., Kelling, S., LeBuhn, G., Littauer, R., Lotts, K., Michener, W., Newman, G., Russell, E., Stevenson, R., & Weltzin, J. (2013). Data Management Guide for Public Participation in Scientific Research. DataONE: Albuquerque, NM.