

Chapter VI: Legal, Ethical, and Policy Considerations
Anne Bowser & Andrea Wiggins

Word Count: 6,155
Number of Figures: 1
Number of Tables: 2

Introduction: The relationship between law, ethics, and policy

When a volunteer is injured through participation in a citizen science project, who is liable for the cost of their recovery? How should ideals of sharing scientific data be balanced with the need for proprietary use, like through publication? And how should these ideals be balanced with the need to protect volunteer privacy? This chapter addresses these questions, and other important considerations related to law, ethics, and policy.

Law, ethics, and policy are three types of guidelines that describe and prescribe interactions between different social agents. Of these guidelines *laws* are most stable, and are enforceable by government entities. For example, The Children's Online Privacy Protection Act (COPPA) is enforced by the Federal Trade Commission (FTC) in the US. Parties that violate COPPA are subject to civil prosecution by the FTC, and fines of up to \$16,000.

Ethics are general moral guidelines. Of the three types of guidelines we discuss, ethics are least stable. For example, many ethics are determined by the specifics of a particular situation. Asking volunteers to collect soil samples may be ethical if the samples are collected from some public properties, such as the yards of public schools, but not from others, such as protected National Parks. Asking volunteers to collect soil samples on private land may or may not be ethical based on how sensitive data, such as the full name and age of participants who collected samples, is treated.

There are two main types of policy. *Public policy* refers to a collection of formal and informal principles that guide decision-making. Some policy directly reflects law; other policy reflects interpretations of the law (such as cases heard by the Supreme

Court). Some public policies may become law, while other policies remain lesser guidelines. Policy is designed to be more flexible than law.

While public policy dictates government practices, *data policy* codifies and dictates the practices of organizations that relate to data [1]. Specifically, data policies are collections of documents that dictate how one party, such as a citizen science project coordinator, may interact with another, such as a citizen science volunteer. Data policies are written to reflect all the considerations described above: law, ethics, and public policy. Common types of data policy include:

- *Legal policies*, which may describe compliance with existing law, and address liability concerns.
- *Privacy policies*, which deal with how data including Personally Identifiable Information (PII) is collected and stored.
- *Terms of use*, which dictate how a product or service (or, for citizen science, data or collection of data) may be accessed and used.

These policies become meaningful when they are embedded within a *user agreement*, or a legal contract between a project and its volunteers.

Data policies are a crucial but often neglected feature of project design. These policies clarify the legal and ethical responsibilities projects have towards volunteers, such as protecting the privacy of volunteers or mitigating physical and emotional harm. Policies also clarify responsibilities volunteers have towards projects, such as taking full responsibility for injuries gained in the course of service, or promising to adhere to community standards of conduct. Finally, data policies clarify data ownership and management.

Throughout this chapter, when we identify an important law, ethical principle, or public policy, we also offer guidelines for how citizen science projects can write their data policies to demonstrate compliance. We cannot over-emphasize the importance of having formal data policies and a user agreement. In some cases, laws such as COPPA explicitly require this documentation as part of compliance. In other cases, documentation is required for practices to be recognized by a court of law.

With that stated, we understand that our readers are scientists and project coordinators driven by their passion for research or education conducted with the help of citizen science volunteers. To make legal, ethical, and policy considerations more relevant, we use the data management lifecycle to organize our discussion of the key concerns facing citizen science practitioners.

Following this introduction, we briefly introduce the data management lifecycle [12]. We then describe the different legal, ethical, and policy considerations that come into play at various stages of the data management lifecycle:

1. *Project planning*: Understanding ethical human subjects research; determining which legal and policy guidelines are relevant to your project
2. *Data collection and quality management*: Recruitment and enrollment of volunteers; liability; user agreements
3. *Describing Data*: Implications for participant privacy
4. *Preserving data*: Guidelines for data storage, including security
5. *Discovering, integrating, and analyzing data*: Copyright and intellectual property; the ethics of open and proprietary data; integrating your data with other data sets

By tying data policies to the data management lifecycle we hope to illustrate why different types of data policy are important, and how they are inter-related. We also want to emphasize that data policies should be established during a project's planning stages and periodically revisited. So, while policies about sharing data are covered in the "preserve and discover" sections of the data management lifecycle, these need to be in place before data collection begins.

As a final note: we are not legal experts, and this chapter is neither offered nor intended as legal advice. Only you know what your project does and what it needs. In addition, we offer examples primarily from the United States legal paradigm, but similar legal, policy, and ethical constraints apply in other countries.

The Data Management Life Cycle

The Data Management Life Cycle (figure 1) illustrates the typical steps in managing research data from creation through preservation and discovery. The activities involved in data management are non-linear. In an ongoing project, multiple processes may occur at the same time, and considerations relevant to every step are addressed during planning. One of the advantages of using a life cycle model is that it readily supports holistic, long-term, big picture planning that can be helpful when considering policy decisions.

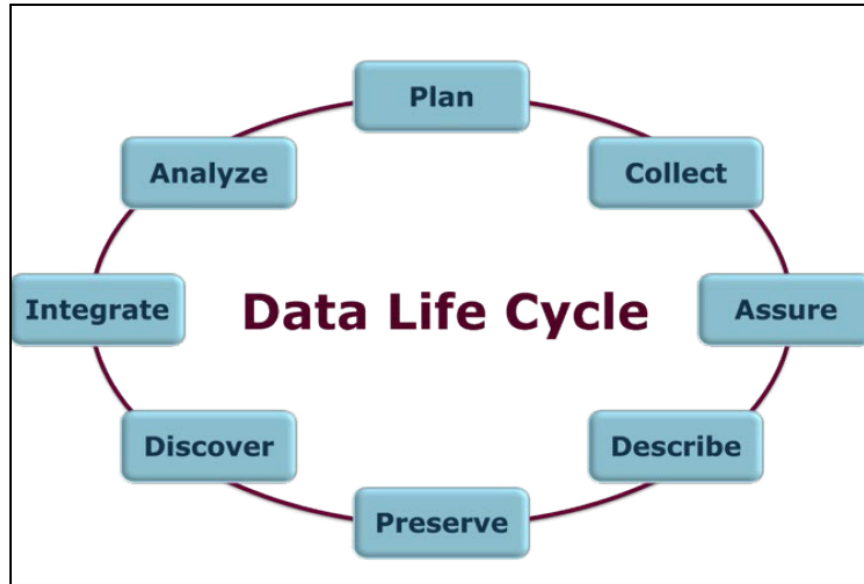


Figure 1: The Data Management Life Cycle [12]

Project Planning

The planning stage of the data life cycle includes mapping out the policies and processes for the entire life cycle, starting with project goals and working backwards to develop policies that meet the project's needs.

One of the first steps is determining the relevant types of legal, ethical, and policy guidelines. Through our work with DataONE and the Wilson Center, we have had the opportunity to speak with a range of citizen science projects about the legal, ethical, and policy guidelines they consider most important to their work. As a general rule, we have observed that similar groups of projects discuss similar types of guidelines. These include ethics approvals, laws specific to federal agencies. We begin by providing a sample of the different legal and policy guidelines relevant to projects of different types, and then present general considerations of the ethics involving the use of human or animal subjects in scientific research.

Legal Considerations

Many projects funded in whole or part by grants from the National Science Foundation (NSF), National Institutes of Health (NIH), and other agencies, are subject to specific guidelines as a condition of the award. Chief among these is the requirement to obtain approval from an Institutional Review Board (IRB) or an animal welfare board such as an Institutional Animal Care and Use Committee (IACUC) to assure the ethical treatment of animals, which can apply to studies of wild vertebrates as well as laboratory animals. Other guidelines—such as those dealing with age-based discrimination, which might impact some projects, and requirements for data management plans—are detailed in grant policies (e.g., the NSF Grant Policy Manual [6]).

Projects run by U.S. federal agencies and projects receiving significant support from federal agencies (other than granting agencies) also express common legal and ethical concerns. For example, these projects must comply with the Privacy Act of 1974, the Paperwork Reduction Act, and the Freedom of Information Act. Projects run by 501(c)(3) non-profit organizations, such as museums and universities, also voice shared concerns (such as adherence with the Volunteer Protection Act). Table 1 presents a list of selected U.S. laws and policies that may impact citizen science projects of these different types.

Note that this is **not** a comprehensive checklist of all laws that may apply to your project, but rather a selection of laws that multiple citizen science projects have identified as relevant to their activities. Projects in other regions are subject to different legal systems and should familiarize themselves with the laws of their own countries that

address the issues targeted by the these US laws and policies. Still other considerations apply to projects operating across international boundaries.

Policy or Law	Affected Parties	Description and Compliance
Children’s Online Privacy Protection Act (COPPA)	Projects run by private sector companies and federal agencies that collect PII from children under 13	<i>Purpose:</i> To protect the privacy and safety of children under 13. <i>Compliance:</i> Post a privacy notice and notify parents of data collection
Privacy Act of 1974	Projects run by federal agencies	<i>Purpose:</i> To protect citizen PII collected by the government. <i>Compliance:</i> Avoid collecting PII; restrict data access to search for non-PII (i.e., a species)
Paperwork Reduction Act (PRA)	Projects run by federal agencies	<i>Purpose:</i> Reduce the public’s information burden <i>Compliance:</i> Complete Office of Management & Budget review ¹
Freedom of Information Act (FOIA)	Projects run by federal agencies	<i>Purpose:</i> Support transparency of government data collection <i>Compliance:</i> Avoid collecting PII; eliminate ability to retrieve records by PII; or, be prepared to share records upon request
Volunteer Protection Act of 1997	Projects run by 501(c)(3) non-profits (most educational institutions) or government agencies	<i>Purpose:</i> To encourage volunteerism by reducing organizational liability <i>Compliance:</i> Have volunteers sign a liability waiver
Institutional Review Board (IRB) approval	Federal grantees; projects run by many federal agencies; projects run by many non-profits	<i>Purpose:</i> Support ethical human-subjects research <i>Compliance:</i> Apply for exemption or for review to an internal IRB third-party service
Institutional Animal Care and Use Committee (IACUC) approval	Federal grantees; projects run by some federal agencies	<i>Purpose:</i> Support ethical research with vertebrates <i>Compliance:</i> Apply for approval

¹ Note that some Federal Projects have found ways around PRA compliance. See, for example, [13].

Table 1: Common laws and policies affecting citizen science projects

For example, nonprofits are exempt from the US-EU Safe Harbor frameworks for online privacy protection, but other projects may be subject to these laws.

Best Practices

After reviewing Table 1, projects are advised to:

- Survey international, national, state, and local laws that pertain to unique aspects of their research
- Survey their own organizational policies
- Verify whether ethics approval is required from an IRB and/or IACUC
- Identify a few citizen science projects that are similar in both the topic studied and the specific methods utilized (e.g. bioblitzes; long-term ecological monitoring), and (if possible) survey their data policies to identify additional considerations

Note that compliance with most laws presented in Table 1 should ideally be embraced as best practices. For example, supporting privacy protections for children along COPPA guidelines introduce some administrative complexity, but can prevent problems and improve public perception of the project. This is an ethical matter as well a practical one. Because laws often expand in their scope or level of scrutiny, it is possible that a legal requirement that does not currently apply to your project now may be judged as applying to your project in the future. Designing policies and supporting administrative processes up front to address as many legal and ethical principles as possible will help reduce the need to redesign policies, or modify or discard data, at a later point in time.

Human Subjects Considerations

In addition to the legal guidelines described above, ethical guidelines for the protection of research subjects should also be considered. In citizen science, the role of volunteers is often difficult to describe. In some ways, volunteers resemble field technicians or scientists, who participate in data collection, analysis, interpretation, and in some cases publication. But citizen science volunteers also resemble research subjects in a variety of ways. The observations they submit may contain data about an organism that *also* contains data about the volunteer (for example, by sharing a volunteer's location). And projects may survey volunteers about their practices and preferences to support general research and individual project design. In both cases, interactions with volunteers are considered human subjects research.

Understanding what constitutes human subjects research requires understanding what is considered research, and who is considered a human subject. In the United States, the Department of Health and Human Service (DHHS) sets the standards for human subjects research [10]. The DHHS defines research as “a systematic investigation, including research development, testing and evaluation designed to develop or contribute to generalizable knowledge.” Investigations designed to support operations are excluded from this definition.

In other words, a project manager who tweaks a single aspect of project design—such as asking volunteers to perform activities alone versus in pairs—is not necessarily conducting research. But if that project manager then publishes an article describing the impact of socialization on data quality, the work becomes research. Many (if not most)

surveys of project participants that are intended to support project operations also have potential to yield research results, particularly if the data may be shared.

Human subjects are defined as living individuals about whom an investigator obtains “data through intervention or interaction...or identifiable private information.” For ecological citizen science, the most relevant portion of this clause is “interaction”, which designates any form of communication or contract between a researcher and a volunteer. This includes in-person interaction as well as online interactions, such as asking a volunteer to take a survey about their project experiences.

Note that most universities provide information on their websites to help researchers determine whether their project is considered human subjects research, and thus should be subject to IRB approval. Because the definition of human subjects research is so ambiguous, projects who aren't sure whether their activities fall under this umbrella are encouraged to consult these resources.

From an ethical perspective, all projects conducting human subjects research should adhere to the national guidelines. From a legal perspective, all projects receiving federal funding are **required** to comply. Citizen science projects operated by some, but not all, federal agencies must comply with these regulations.² Projects run by non-profits should consult with their specific institutions (noting that most universities require IRB approval and provide extensive supporting resources).

Generally, human subjects research compliance requires three main categories of actions:

1. Submission of the proposed research for IRB review

² For a list of agencies that must comply, visit <http://www.hhs.gov/ohrp/humansubjects/commonrule/>

2. Ensuring that volunteers have opportunity to give informed consent before participating in human subjects research
3. Providing a publicly-accessible statement of compliance to pertinent regulations

Ethical guidelines for conducting human subjects research are provided by a range of international organizations, such as the United Nation's World Health Organization and by national authorities such as the DHHS in the US. In some cases, guidelines also address research conducted in specific domains like medicine, though unfortunately no guidelines currently exist for citizen science researchers and practitioners.

In the absence of specific guidelines, we believe that the principles published in the Belmont Report³ can be particularly valuable for citizen science projects. The Belmont report identifies three basic ethical principles, "generally accepted in our cultural tradition" and particularly relevant to research ([5], p.4):

1. *Respect for persons* suggests that individuals are autonomous agents deserving protection from activities that might harm them
2. *Beneficence* suggests that steps should be taken not only to prevent individuals from harm, but also to actively secure their well-being
3. *Justice* suggests fairness in distribution of research benefits and risks

These guidelines are enacted by a number of citizen science projects in various ways. In accordance with *respect for persons*, many projects provide volunteers with online or in person training to support safety through information and demonstrated

³ The Belmont Report was first published in 1978 by researchers working for the Department of Health and Human Services. The Report was commissioned by DHHS following controversy over the Tuskegee Syphilis Study, where a disadvantaged population of African-American sharecroppers were deliberately infected with Syphilis and left untreated for purposes of research. The Belmont report is considered an ethical standard within the United States and informs codes published by DHHS and other federal agencies, including the National Science Foundation (NSF) and the Office of Human Research Protections (OHRP).

practices. Supporting participant autonomy can also involve, for example, allowing volunteers to choose where they participate in data collection activities.

Projects implement *beneficence* in ways that reflect their unique goals. Project Budburst, a plant phenology project, is designed to support science and also formal and informal education. To the later, Project Budburst offers low-cost online courses with over 30 hours of material through the Citizen Science Academy.

And the Cornell Lab of Ornithology's Celebrate Urban Birds project embodies the ethical principle of *justice* by targeting historically underserved Latino communities, thus ensuring that the benefits of participating in citizen science are extended to diverse social groups. Piloting new tools and procedures is another area where justice may apply: if testing shows that participant experiences are meaningfully improved by an experimental treatment, *justice* suggests that all volunteers should be granted access to these means.

Best Practices

- Determine whether formal IRB assessment is necessary based on the potential for producing *general* knowledge (such as that intended for academic publication)
- Determine whether IACUC assessment is necessary based on the level of involvement with non-human vertebrae
- Consider the ethical principles codified through the Belmont report; develop a short list of ethical principles to be shared internally or published online

Data Collection and Quality Management

Data collection guidelines usually focus on the design of procedures or methods for ensuring data quality. But from a policy perspective, the primary considerations are instead the liability considerations that arise when engaging volunteers in data collection,

and policies related to the use of technologies for reporting data. User agreements between projects and volunteers codify these considerations, along with data use.

In U.S. law, liability is a legal obligation to pay debts or damages. In citizen science, tort liability—which occurs when the actions of party cause damages to another party, who then files a key complaint—is a key concern of both projects and volunteers. If a volunteer becomes injured in the process of collecting data, the project leaders or the organization hosting the project could be faced with a lawsuit. Generally, projects manage this risk is to have volunteers sign a waiver of liability.

Researchers who study liability in sports and recreation (a similar context of voluntary participation) identify four factors that increase the likelihood a liability waiver will hold up in court [3].

- Waivers should contain clear and unambiguous language.
- Waivers should not codify expectations that breach law or policy, for example by asking volunteers to collect data on private land.
- Waivers must be presented in the form of a contract that both parties explicitly agree to. In other words, waivers hosted on static web pages are unlikely to hold up to scrutiny, which is problematic for many citizen science projects.
- Waivers must be signed by parties who are legally able to contract (i.e., healthy adults over the age of 18).

Note that the above are general guidelines for writing waives to protect projects run in conjunction with formal authorities, such as a non-profit, government agency, or academic institution. Some researchers suggest that projects run out of federal agencies may generally be protected from liability through precedent, like through court decisions

related to the Federal Torts Claim Act [9]. Projects that are self-organized and run by volunteers may find additional guidance in [8].

Data quality assurance processes, like data policies, are active at multiple stages of research. To a large extent, data quality is dependent on the participation protocols and tools used to submit data. The needs for additional detail about observation data in particular must be weighed against the potential risks of acquiring it.

When deciding which types of data to collect, we encourage project coordinators to ask the following questions:

1. What information is necessary...
 - a. To answer research questions or otherwise achieve project goals?
 - b. To address similar research questions or support related goals?
2. Are there risks associated **with the type of information collected**?
 - a. Is the process of data collection sensitive (i.e. may disturb vulnerable habitats; data is collected on private land)?
 - b. Is the data itself sensitive (i.e. contains information on the location of sensitive species)?
3. What tools and technologies will be used to collect, upload, and share data?
 - a. What are the risks associated with these tools (i.e., most Smartphones automatically collect sensitive information on volunteer location)

Answering these questions often requires balancing different goals: for example, a project may be required to choose between collecting extra data types to support future research, or limiting collection to protect volunteer privacy. Coordinators should also note that volunteers may unintentionally provide personal information, like by including people in

photographs of species. While there are no right or wrong answers to the above questions, project coordinators have the ethical responsibility to make informed decisions and also to share these decisions with their volunteers. These decisions are communicated through data collection protocols, but also (and, from a legal perspective, more importantly) through the implementation of a user agreement.

A user agreement is a contract between an organization (here, a citizen science project) and an employee, volunteer, or another organization. There are two main types of user agreements. *Clickwrap* agreements are explicit agreements between a volunteer and a project. These agreements take the form of paper forms signed by volunteers, or online forms where volunteers click a box saying “I agree” to a set of terms and conditions. In contrast, *Browsewrap* agreements are implicit agreements. These take the form of posted policies that volunteers may see, but do not directly agree to abide to.

Legal review of these two agreement types suggests that clickwrap agreements are significantly more likely than browsewrap agreements to hold up in a court of law [7]. Key components may be providing users with *adequate notice* and *meaningful opportunity* to review terms and conditions. For example, one court ruled that the act of clicking “continue” to move past an online form stating “I have read and agreed to [provider] terms and conditions,” **and linked to those terms and conditions**, constituted a valid user agreement.

Best Practices

- Using the guidelines above, construct a liability waiver describing the potential risks of participation, and stating that any damages incurred by volunteers, or

caused by volunteers through participation, are the full responsibility of those volunteers

- Consider exactly which types of data you wish to collect, and communicate this decision with volunteers through protocols and through also terms of use
- Construct a clickwrap user agreement linking to information about participation including a liability waiver, a description of data collection activities, a data use policy (described later in the chapter), and privacy policy (described directly below)

If possible, projects are also encouraged to seek legal council during the initial stages of planning and protocol design.

Describing Data: Implications for Participant Privacy

Data description requires documenting the data set and specifying metadata that describe the data. Data description is relevant at two levels in citizen science projects: the project and the data record. First, at the project level, metadata describes the aggregate data set generated by volunteers and provides documentation of the research procedures and column headers for exported database files. This documentation is provided so other researchers and volunteers can understand, generally, what types of data are collected.

Project-level data documentation rarely poses risks or concerns for volunteers.

Second, at the record level, metadata about individual measurements or observations is important to analysis and data sharing, and may include effort information that is also indirectly “about” the volunteer who contributed the record. Record-level data documentation therefore has potential to reveal details about volunteers’ identities and must be handled with care [2]. For NSF grantees, protecting volunteer privacy is a policy

consideration: according to the NSF Grant Policy Manual, “Privileged or confidential information should be released only in a form that protects the privacy of individuals and subjects involved (p. 105).” Protecting volunteer privacy is also an ethical best practice, and reflects the principle of Respect for Persons codified in the Belmont Report.

There are no universal solutions to the potential conflict between providing maximal privacy protection for volunteers and the scientific ideals for information on effort, location, and other details that could put volunteer privacy at risk. In cases where privacy threat is a realistic scenario, it is important to make a considered choice and document it in project policies. Even within a project team, data sharing requires making policy decisions regarding how to handle participant privacy; other considerations around the decision to share data are discussed in depth later. The privacy policies presented on most citizen science project websites primarily cover the website itself (like by describing whether cookies are used) and do not always mention how volunteers’ privacy will be safeguarded when project data is used for scientific research and related (often secondary) conservation, recreation, and educational purposes.

There are various guidelines for identifying which aspects of data are most important to de-identify or anonymize. One particularly valuable set is the Health Insurance Portability and Accountability Act (HIPPA)’s “Safe Harbor” method⁴ for the de-identification of medical data. These guidelines focus on the removal of 18 unique identifiers, and have been widely embraced within and outside of the medical domain.

Identifiers relevant to ecological citizen science include:

⁴ Not to be confused with the Safe Harbor rules for complying with privacy standards of the European Union (EU), which were developed in collaboration between the U.S. Department of Commerce and the EU.

- Name
- Social security number
- All geographic subdivisions smaller than state
- All elements of dates (except year), for dates related to an individual
- Telephone number, fax number, email addresses
- IP Addresses
- Device identifiers and serial numbers for mobile technologies
- Certificate/ license number
- Full face photographs and comparable images

Of course, we do not suggest that your project should never collect PII or other sensitive data types like those discussed above. Rather, it is important to be aware of the potential sensitivities associated with the data you collect, especially when setting up operating procedures and selecting tools or repositories for storing your data, or deciding whether and how to share your data with volunteers, other researchers, and the public at large.

To this end, researchers studying privacy have identified how different types of data correspond to different levels of privacy risk. A synergy of contributions from [4] and [11], applied to the domain of ecological citizen science, is presented in Table 2.

<i>Level</i>	<i>Description</i>	<i>Example</i>	<i>Risk</i>
1: Raw data	Clearly identifiable data	Participant name, physical mailing address, email address	High
2: Masked data	Use of pseudonyms	Pseudonym replaces name; no other data is altered	High/ medium
3: Exposed data	Masking key attributes	Changing birth date to birth year; reducing GIS location precision to 1 km radius	High/ medium
4: Managed data	Obfuscating key attributes and objective evaluation of risk	Changing birth date to birth year, in accordance with HIPPA and US-EU Safe Harbor guidelines	Low/ medium
5: Aggregated/ anonymized data	Data contains no potential identifiers	Paper checklists of species occurrence data taken in a public park and returned to a drop box	Low

Table 2: Data Sensitivities

As noted in the section above, it is important to include information on which types of data are collected, and how these data will be used and shared, in data policies. This information might be placed on a page labeled terms of use, or on a separate “privacy policy” page.

Best Practices

- Consider exactly which types of data you wish to collect, and evaluate how this data may threaten the privacy or security of volunteers
- If possible, decide which types of sensitive data you will anonymize, and how
- Communicate this information to your volunteers through data policies, such as terms of use or a separate privacy policy

Preserving Data

Like documenting data, preserving data serves multiple interests. The most obvious among these is supporting the scientific ideal of repeatable research; a less obvious but no less important reason is supporting the expectation of volunteers that you

will safeguard the data they invested their time and energy to contribute. Data preservation focuses on avoiding data loss in the short term, and on ensuring access to data in the long term.

Short-term data preservation policies are more likely to take the form of standard operating procedures for data backup and secure storage. While there are no widely known instances of citizen science projects encountering data security problems, there is also no special immunity to viruses or identity thieves. Whenever possible, online data submission should be conducted through secure web forms to protect participants' privacy.

Long-term data preservation involves additional policy considerations: will data be deposited in a repository, and if so are there any limitations that should be placed on access? Wherever reasonable, making data "open access" (freely available online) is increasingly required for publicly funded research, and is often in keeping with the goals of citizen science projects. If data updates are regularly deposited, how will retroactive changes to data (editing or deleting data points) be handled? In most cases, "deleted" accounts or data are not removed from data that have been already archived for both practical and scientific reasons, but are excluded from additional data updates moving forward.

Best Practices

- Decide how to best preserve your data in the short- and long- term to balance security needs with ideals of re-use
- Communicate this information to your volunteers through data policies, such as terms of use

Discovering, Integrating, and Analyzing Data

Citizen science projects collect data to achieve a number of research, education, and policy goals. Based on these goals, projects may decide to share different levels of data, with different parties, and at different times. This section begins first by considering data ownership, and then by exploring how parties may share their data, or integrate their data with other data sets. We conclude with a quick nod to the current debate around the ethics of data sharing versus proprietary use.

In some cases, data collected by citizen science projects is a form of intellectual property subject to copyright. *Intellectual* property is a legal term that protects the rights of “creations of the mind,” including patents, creative works, and trade secrets. *Copyright* is a legal term that designates ownership of creative works. Examples of creative works include individual efforts such as novels or paintings, and also *creative compilations*.

According to The U.S. Copyright Act, creative compilations are produced through "collection and assembling of preexisting materials or of data that are selected in such a way that the resulting work as a whole constitutes an original work of authorship." Copyright law is important for citizen science practitioners to understand because it relates to the ownership of compiled data sets, such as those collected by scientists with the aid of volunteers.

The 2001 Supreme Court *Case Feist Publications, Inc. v. Rural Telephone Service Company, Inc* clarifies the copyright limitations of such creative compellations. This case reached the Supreme Court when a small, rural publisher of a telephone directory (Rural Telephone Service Company) refused to share their directory with a larger, regional publisher (Feist Publications). The Supreme Court ruled that Rural

Telephone Service Company could not claim copyright of their directory for two reasons. Neither their method of data *selection*—namely, including all residents in a particular geography—nor their method of data presentation—intuitive alphabetical listing—involved sufficient originality to merit copyright protection. *Feist* also clarifies that the factual data within a database (for example, the street address of a particular individual) never enjoys copyright protection.

There are currently no legal cases that we know of addressing whether citizen science data sets can be considered creative compilations. But, based on the *Feist* ruling (and repeating that this chapter does not constitute legal advice) we extrapolate that these data sets probably are creative compilations. If so, they would be subject to copyright law. All forms of aggregate data publication—ranging from newsletters shared with volunteers, to academic publications—are creative contributions subject to copyright.

In the United States, copyright is automatically established when a new creative work is produced. As a right, copyright gives the creator of a work exclusive right to use and distribution for a finite period of time (currently, the author's life plus 70 years). Citizen science projects that wish to share their data can do so through two primary mechanisms: licensing, and waiving copyright.

A license grants formal permission for one party to access the intellectual property of another, and dictates the acceptable terms of use. For example, a project might permit the re-use of its data for research and education, but not for commercial use. Or, a project might require that a copyright notice linking to the project's web page accompanies all data that is reused. While these licenses are often posted to a project's

data policies page, we emphasize that they are most meaningful when embedded in a clickwrap user agreement.

The Creative Commons produces one collection of licenses designed to support data sharing under various conditions. These are posted to the Creative Commons website, at <http://creativecommons.org>. The Creative Commons website also hosts an online tool to generate customized licenses including title of work (e.g. data set), name (e.g. of project), and url.

In rare cases, a project may wish to completely waive ownership of their data. Designating a data set as part of the public domain may achieve this end. However, projects that put their data in the public domain may still be under certain obligations, such as the moral right to object to abuse of the data (and the Creative Commons recommends against placing copyrighted data in the public domain for precisely this reason).

Project data policies should assert ownership of the data sets collected, and also describe conditions for reuse. This information supports volunteer agency in deciding whether to contribute to a project, and also helps other researchers discover and use data sets. It is equally important for projects to consider the stated policies of other data sets included in research. For example, while most government data (such as census data) is designated as “open” for all to access and re-use, data collected through private sector institutions may place restrictions on re-use. Some citizen science projects simply require that data sets are properly cited. Other projects ask that a project lead is included as a co-author on any publication resulting from data use. These policies should be adhered to as best practices from an ethical point of view, and may also be legally binding.

We do not mean to suggest that all projects should share data through the mechanisms described above. We recognize that the level that each project can and should share data, both with other scientists and with the general public, is highly contested. On one hand, data sharing promotes new knowledge and may be a best or even required practice. For example, the NSF suggests, “Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants [6].”

On the other hand, many researchers rely on publication to support their careers. In such cases, the success of current and future citizen science projects may depend raw data private, at least until the results of a study can be published. Researchers should also resist sharing data sets when doing so could put participant privacy at risk—a point also acknowledged by the National Science Foundation in their Data Policy Manual.

Best Practices

If you believe that your data set is a form of intellectual property:

- Decide how and whether you want volunteers, external researchers, and the general public to access your raw data
- Communicate this information to your volunteers through data policies, such as terms of use

If you will integrate your data with external sources:

- Carefully review the associated data policies, including terms of use

Conclusion

This chapter describes the legal, policy, and ethical considerations associated with ecological citizen science. In addition to describing these considerations, we present best practices for designing policies that relate to each consideration raised. Note that it is insufficient to simply enact these policies through practice; in order to serve legal and ethical goals, practices must be formalized in data policies explicitly agreed through in a user agreement.

We re-iterate that the authors are *not* legal experts, and the information in this chapter is not legal advice. The authors *are* researchers (and occasionally practitioners and volunteers) committed to serving the needs of the citizen science community. To this end, we encourage our readers both to consult legal professionals for legal advice, and to contact us with any questions that we might answer in our limited, but compassionate, capacity.

Literature Cited

- [1] Bowser, A., Wiggins, A. and R. D. Stevenson. 2013. Data policies for Public Participation in Scientific Research: A primer. DataONE. Albuquerque, N.M. 13 pp.
- [2] Bowser, A., Wiggins, A., Shanley, L., Preece, J. and S. Henderson. 2014. Sharing data while protecting privacy in citizen science. *ACM Interactions* XXI.1: 70-73.
- [3] Cotten, D. and M. Cotton. 1997. Legal aspects of waivers in sports, recreation, and fitness. PRC Publishing, Canton, OH.
- [4] El Emam, K. 2010. Risk-based de-identification of health data. *IEEE Security Privacy* 8: 64-67.
- [5] The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report. Retrieved from http://videocast.nih.gov/pdf/ohrp_appendix_belmont_report_vol_2.pdf
- [6] National Science Foundation. Proposal and Award Policies and Procedures Guide, February 2014. Retrieved from http://www.nsf.gov/publications/pub_summ.jsp?ods_key=papp
- [7] Organ, S. and M. Corcoran. 2008. Your web site's "Terms of Use": Are they enforceable? *Privacy & Data Security Law Journal* 3: 110-114.
- [8] Robson, E. 2012. Responding to Liability: Evaluating and Reducing Tort Liability for Digital Volunteers. Woodrow Wilson International Center for Scholars. Washington, D.C. 68 pp.

- [9] Smith, B. 2014. Agency liability stemming from citizen-generated data. Woodrow Wilson International Center for Scholars. Washington, D.C. 8 pp.
- [10] U.S. Department of Health and Human Services. Human Subjects Research (CRF 46). Retrieved from <http://www.hhs.gov/ohrp/humansubjects/guidance/>.
- [11] U.S. Department of Health and Human Services. Guidance regarding mobile methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Retrieved from http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveridentities/De-identification/guidance.html#_edn1
- [12] Wiggins, A., Bonney, R., Graham, E., Henderson, S., Kellig, S. LeBuhn, G., Littauer, R., Lotts, K., Michener, W., Newman, G., Russell, E., Stevenson, R. and J. Weltzin. 2013. Data management guide for public participation in scientific research. DataONE. Albuquerque, N.M. 37 pp.
- [13] Young, J., Wald, D. Earle, P. & L. Shanley. 2013. Transforming Earthquake Detection and Science Through Citizen Seismology. Woodrow Wilson International Center for Scholars. Washington, D.C. 64 pp.