



Combining Physical and Digital Data Collection for Citizen Science Climate Research

HEATHER KILLEN 

LUCY CHANG 

LAURA SOUL 

RICHARD BARCLAY 

**Author affiliations can be found in the back matter of this article*

CASE STUDIES

 ubiquity press

ABSTRACT

In this paper, we present our experience designing and implementing a hybrid citizen science protocol combining local data collection reported digitally with the return of physical samples by mail. Our project, Fossil Atmospheres, housed within the Paleobiology Department of the Smithsonian Institution's National Museum of Natural History, sought to complete a broad geographic collection of *Ginkgo biloba* L. leaves to better understand climate change over time. We also wished to leverage and test the affordances of using an established online platform as a technological tool for research-quality data collection. Participants were asked to find a local ginkgo tree and, using a hybrid protocol, collect leaf samples and record site data, including photos, GPS coordinates, and tree characteristics, using the iNaturalist online platform. Participants then returned their leaf samples by mail. Fossil Atmospheres received 562 leaf samples from 352 participants. These samples, representing 36 states, met our target geographic transects and reflected the known habitat range of living ginkgo in the United States. We were able to successfully pair a large majority of received samples to their corresponding digital data records, allowing us to include 88% of the samples received within the Fossil Atmospheres data set. These results greatly exceeded our project goals. The hybrid protocol model we present, based on our experiences, indicates that using tools like iNaturalist provides multiple benefits that meet or exceed more traditional data collection models, including increases in the scale of data that can be collected, data accuracy, and data completeness, uniformity, usability, and accessibility.

CORRESPONDING AUTHOR:

Heather Killen

University of Maryland, College Park, US

heatherakillen@gmail.com

KEYWORDS:

citizen science; technology; physical sample; digital data; iNaturalist; protocol design

TO CITE THIS ARTICLE:

Killen, H, Chang, L, Soul, L and Barclay, R. 2022. Combining Physical and Digital Data Collection for Citizen Science Climate Research. *Citizen Science: Theory and Practice*, 7(1): 10, pp. 1–14. DOI: <https://doi.org/10.5334/cstp.422>

INTRODUCTION

Citizen science has long played a role in advancing scientific knowledge, especially in projects that require data spanning large geographic or temporal ranges (Bonney et al. 2016). Technology has become a central component of citizen science (Newman et al. 2012) as a tool for engagement (Aristeidou et al. 2017), community building (Peterman et al. 2019), and data collection (Wittmann et al. 2019). The experiences gained from novel implementations of technology in citizen science are thus broadly applicable. Paleontology has frequently engaged people outside of academia to assist with the work of collecting and documenting fossil specimens, allowing for larger-scale excavation than would otherwise have been possible. Recent efforts to engage this amateur community with paleontology online in a way that can support research have seen success (e.g., myFOSSIL, n.d.). Here we build upon the existing success of these projects and report on our experience using an established online platform as a technological tool for data collection and storage together with physical sample collection.

One model for citizen science projects involves asking for help to collect local physical samples (Pandya 2012). Another is for projects to collect local observations, rather than physical samples, that are then reported virtually (MacPhail and Colla 2020). Still other projects are completed completely online, such as those housed within the Zooniverse platform (Simpson et al. 2014). An example of this third category is our Fossil Atmospheres stomatal index effort, where participants are asked to count features on microscopic imagery (Soul et al. 2019). The work we present here represents a fourth type of citizen science model that links national sample collection with virtually reported data. In this effort we join other successful ecological projects, such as the fungal diversity project FunDiS (Sheehan et al. 2021).

In the summer of 2019, we asked citizen scientists across the country to collect both observational data, which was reported digitally, and associated physical samples that were mailed to us. This required navigating four primary design challenges: 1) developing a clear, concise, two-part, hybrid protocol using technology that facilitated easy and correct collection of both physical samples and associated site data by citizen scientists; 2) composing clear, concise instructions regarding how to complete this protocol in the absence of any direct contact between citizen scientists and the research team; 3) creating a user-friendly vehicle, in our case a website, to virtually deliver the protocol and instructions to participants and; 4) designing robust data practices, including a data management plan and sample receipt protocol, that allowed us to pair physical samples

to their virtual data. By reporting our experience, we hope to provide valuable insights that will be useful to future projects as these technological tools continue to develop.

THE PROJECT

Fossil Atmospheres is a large climate science research project housed in the Department of Paleobiology within the Smithsonian Institution's National Museum of Natural History (NMNH) in Washington DC, USA. Our research uses fossil and modern ginkgo leaves to develop baselines to understand ancient and future climate change (**Figure 1**; Barclay and Wing 2016). Specifically, we use the leaf stomatal index, a measure of the percentage of stomatal cells in relation to all cells on a leaf surface, as a climate proxy (Woodward 1987). Ginkgo plants are an excellent study subject as the Ginkgoaceae family has a 200-million-year plus fossil record. Today Ginkgoaceae is represented by a single surviving species, *Ginkgo biloba* L. (henceforth referred to as ginkgo). Ginkgo is a ubiquitous modern landscape tree that is broadly recognizable by the public due to its unique leaf shape and popularized for its purported medicinal uses.

The Fossil Atmospheres project is made up of three initiatives that each contribute a unique source of leaves from which we calculate stomatal indices. The first source of leaves comes from herbarium specimens and fossils dating back to 56 million years ago, which are used to establish historic baselines for the index. The second source consists of leaves from trees grown under experimental CO₂ conditions. The third source, which is the focus of this paper, consists of modern ginkgo leaves from across the full geographic range of ginkgo in the United States to determine how naturally varying the stomatal index is today.

To achieve this third initiative, our team, composed of paleobiologists, science educators, and museum volunteers, developed a hybrid citizen science protocol that asked people throughout the United States to complete two connected tasks: collect and submit physical ginkgo leaf samples, mailed to us at the NMNH, and collect tree images, and location and other site data reported using the iNaturalist platform. Upon receipt, all leaves were accessioned into the permanent Smithsonian collection. By launching a nationwide call for ginkgo leaf samples, we hoped to capture a snapshot of the species across a wide geographic range. As long-term repositories of physical samples and information, museums are an ideal venue for such projects. We anticipate future researchers will access the collection to answer new research questions beyond our investigations, for example those involving genomic sequencing, particularly as new technologies for analysis emerge.



Figure 1 Image of fresh, herbarium, and fossil ginkgo leaves. Fresh leaves are from the Fossil Atmospheres experiment at the Smithsonian Environmental Research Center. The dried leaf represents herbarium material collected during the past century. The fossil is from the late Paleocene-aged Almont locality in North Dakota (~58 myo). Image used with permission from Scott Wing.

TECHNOLOGY AS A TOOL FOR COLLECTING CITIZEN SCIENCE DATA

Citizen science projects regularly ask for data to be uploaded to a project-specific website or mobile application. Numerous citizen science projects have also been exploring how increasingly popular online platforms, such as eBird or iNaturalist, might support citizen science. Projects are using online platforms to collect robust scientific data (Sullivan et al. 2014) but also to engage, support, and include a community in scientific or policy decisions (Groom et al. 2019). One powerful affordance of online platforms can be that they provide both machine learning algorithms and a community of knowledgeable users that allow for quick and accurate confirmations of the identity of an organism (Unger et al. 2020). Accurate identification has been shown to be a successful tool for projects that are interested in mapping species habitat (La Sorte and Somveille 2020), and for documenting species richness within an area for either specific research projects (Wittmann et al. 2019) or general BioBlitz efforts (Parker et al. 2018).

A second affordance of online platforms is the standardization of data collection protocols, which mitigates a traditional concern within citizen science

regarding varying data quality. Using an online form for data entry is one way to control for incomplete or missing data—recognizing that best practice encourages flexibility with data collection tools so that possible participants or locations are not excluded because of limited technology access (US GSA, n.d.). Established online platforms, such as iNaturalist, may also allow individual projects to incorporate customizable data fields beyond those found in a standard record. This flexibility allows timestamped documentation of ecological and environmental conditions associated with a particular sample, adding to the depth and robustness of data associated with it.

Associating digital data, citizen science, and natural history museums allows for additional benefits. Museum specimens can be used to further validate digital records data and vice versa (Spellman and Mulder 2016), providing a strong case for the use of museum and online records in combination, as we have done in the Fossil Atmospheres project. Combining museum collections with iNaturalist records makes best use of the strengths of each of these repositories to result in a long-term physical asset with rich associated date- and location-specific metadata.

METHODS

Project development involved three distinct phases: 1) choosing an online platform and a sampling period; 2) designing robust data practices; clear, concise sampling protocols; clear, concise instructional guides; and an effective user interface; and 3) engaging with community partners and social media to drive participation (*Figure 2*).

CHOOSE

We began by choosing an online platform that would serve as the best digital tool for both our project and our sampling period.

Selecting an online platform

There are several options for digital collection and repository tools. Developing our own data submission platform through a custom website, mobile application, or online form would have allowed for maximum flexibility when designing workflow and user experiences. However, de novo platforms fail to leverage pre-existing communities of citizen scientists, and the development and maintenance of project-specific platforms can be resource intensive and difficult to keep updated for the newest operating systems. By instead choosing an established platform with long-term institutional support, we took advantage of a tested and reliable data collection tool optimized for mobile and

web use, while benefiting from an active, pre-existing user community.

We considered two main platforms, iNaturalist and *CitSci.org*. Both offered similar functionality and robust participant engagement. After consideration, we decided the iNaturalist platform was more appropriate for the Fossil Atmospheres project because of its large, pre-existing community of contributors, its user-friendly mobile and web interface, and its well-developed functionality to self-build projects. iNaturalist (*iNaturalist.org*) is an “online social network of people sharing biodiversity information to help each other learn about nature . . . [a] crowdsourced species identification system, and an organism occurrence recording tool” (Seltzer, 2021. para. 1) that allows users to submit and curate observations of the flora and fauna they encounter. It is a joint initiative by the California Academy of Sciences and the National Geographic Society with almost 90 million observations by more than 2 million users as of February 2022 (iNaturalist, n.d.).

At minimum, each iNaturalist observation requires a photograph, location, and species identification. Species identification is scaffolded through computer vision classification. Immediately after a photo is uploaded, the iNaturalist algorithm presents the user with a choice of likely species. The user’s choice is then confirmed by human users of the platform. iNaturalist is well known for having scientific experts as users and for producing accurate

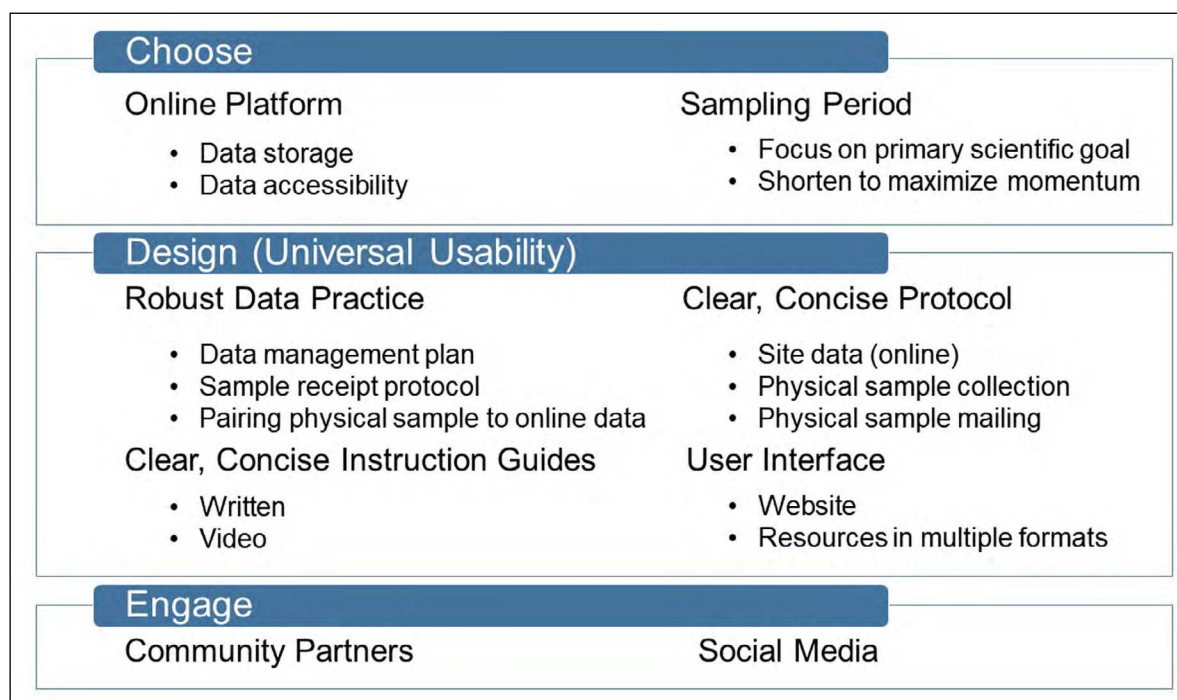


Figure 2 Three phases of project development and implementation when utilizing a hybrid protocol: Choose, Design, and Engage. Phases are presented in implementation order (*top to bottom*) and include subordinate categories representing areas for action with primary goals.

identifications. This species recognition was a powerful feature when observing ginkgo, as the unique leaf shape was easily recognized by computer vision.

iNaturalist also allowed for the creation of projects on the platform. Creating a project within iNaturalist allowed our team to save, track, and display observations in real time. It also allowed project administrators to communicate with project members by publishing comments, journal posts, or guides. At the time of our project setup, iNaturalist allowed project organizers to create required and optional custom data fields beyond those required by a typical iNaturalist observation. In this way, we were able to ensure that each observation submitted to the Fossil Atmospheres iNaturalist project included the data critical for our research study, such as a ginkgo tree's height and sex. Screenshots of the Fossil Atmospheres iNaturalist project page have been included in the supplemental material (Appendix A).

Selecting the sampling period

Determining the sampling period was primarily driven by our project's scientific considerations. Ginkgo is unusual in that it has male and female trees, which may respond differently to surrounding CO₂ levels. We therefore needed to have citizen scientists determine and record tree sex. In August, female trees have noticeable fruit and most leaves are fully matured but have not yet begun to yellow. We therefore chose August as our primary sampling period. We also hoped a one-month collection period would act to maximize collecting momentum.

DESIGN FOR UNIVERSAL USABILITY

We needed a protocol that facilitated easy and accurate collection of both physical samples and associated site data that could subsequently be paired by the research team. Because we expected citizen participation to be self-guided, or perhaps facilitated by staff at botanical gardens and herbaria but never directed by project researchers, we focused on designing a collection protocol that employed materials and platforms that were widely available and user-friendly. In this way, we hoped to lower the barrier to participation while acquiring quality specimens and associated digital data.

Developing a clear, concise sampling protocol

Our sampling protocol involved two tasks: 1) collecting observational data at the tree, recorded digitally and 2) collecting a sample of leaves at the tree to mail to the NMNH. The complete hybrid protocol is included in the supplemental materials (Appendix B). To develop the collection protocol, we focused on identifying single steps and data fields that would allow us to extract multiple pieces of information. For example, team scientists were

interested in aspects of the tree's habitat and specific sampling details. Rather than asking participants to record that information, we instead asked for photos of the base of the tree. The research team could later use these photos to determine that information. In this way, our protocol shifted the data processing from citizen scientists to our team as much as possible.

When a participant submitted their observation to the Fossil Atmospheres iNaturalist project, they included two photos, the geo-tagged location and date of the observation, a positive identification of the pictured tree as ginkgo, and three custom data fields specific to our project. All custom fields provided options from a drop-down menu, ensuring data standardization (Appendix C). We first asked participants to estimate the height of the tree from two defined ranges (10–30 feet or 30+ feet). We then asked that participants determine the sex of the tree (male or female). Because iNaturalist requires only one photograph for an observation to be submitted to their platform, our third field prompted participants to confirm that they had uploaded two photos to meet protocol requirements: of the tree and of the base. Finally, we asked participants to determine, if possible, which side of the tree the leaves were collected from (north, south, east, west, or unknown). This was not required, as we did not want to preclude or discourage any participants who may not have had the technology or the skills to easily collect this information.

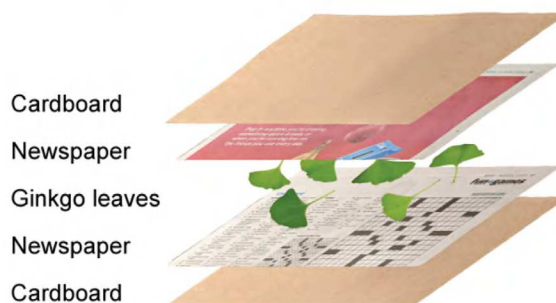
The protocol for submitting the leaf samples was designed to minimize the risk of physical damage during shipping or biological damage from factors such as mold if mailed packages were not received in a timely manner. We also wanted a package that could be mailed cheaply given the United States Postal Service's parcel size thresholds. We focused on using packaging materials that were easily accessible around a home or office, eventually deciding on cardboard, tape, and newspaper. The shipment protocol was iteratively tested and refined by mailing leaves to ourselves at the intended destination address at the NMNH. We found that making a "cardboard sandwich" (Figure 3), with the sample leaves surrounded first by newspaper and then by cardboard, all of which was secured by tape and then placed in an envelope for mailing, was the minimal amount of effort required to adequately protect the leaves. International citizen scientists were asked to dry specimens thoroughly before shipment and provide extra documentation to comply with government requirements. All samples, including those sent internationally, were received in good condition.

To ensure an adequate amount of sampling, participants were asked to collect at least six leaves from a single shoot on a tree. Critically, each sample of six leaves was to be contained within its own cardboard

Instructions

Making the cardboard sandwich

1. If you are sending more than one set of leaves in the same shipment - for example, if you collected from multiple trees or were part of a group that made multiple collections, each collection must be in its own sandwich, so please repeat steps 4B through 4F for each set of six+ leaves, then put all the sandwiches in the same envelope.
2. Each set should have its own unique observation already submitted to iNaturalist.
3. Sandwich the collected leaves between two or more sheets of newspaper. Leaves should lie completely flat and not bent or overlapping.
4. Place two sheets of rigid cardboard around the newspaper layers. In the end, your cardboard sandwich should look something like this:



5. Wrap tape around all four edges of the cardboard sandwich tightly so that no items within it move during shipping.



6. Write clearly in marker or pen your (1) **iNaturalist username** as well as the (2) **date** and (3) **time of the observation** you submitted in iNaturalist on the outside of each cardboard sandwich. We need to be able to link the leaves we receive to the data you submitted!
7. Put all the cardboard sandwiches in an envelope or package, and **before the end of August** mail it to:

Fossil Atmospheres
 Smithsonian NMNH - Dept. of Paleobiology
 10th and Constitution Ave NW
 Washington, DC 20560

Congratulations! Thank you for partnering with us and contributing to ongoing scientific research.

Figure 3 An excerpt of the protocol instructions detailing how to create the cardboard sandwich used to ensure intact delivery of leaf samples to the Smithsonian. Full protocol instructions are available in the supplemental materials (Appendix B).

sandwich with the exterior cardboard layer clearly marked with the participant's iNaturalist username, date, and time of observation. This information allowed the Fossil Atmospheres team to identify which unique iNaturalist observation the leaves were associated with and precluded the need to collect any additional personal information not already publicly available on iNaturalist's website.

Developing clear, concise instructional materials

Previous research has found that in-person training is a factor in the success of citizen science projects that ask participants to follow a data collection protocol (Kosmala et al. 2016). Given the broad geographic scope of our collection effort, face-to-face training in sample collection was logistically unfeasible. As we did not expect to interact with our participants in person, we devoted considerable time and careful effort to designing clear, concise instructional materials.

Instructional materials consisted of a video introducing the goals and research behind Fossil Atmospheres, an instructional video detailing how to complete the hybrid protocol, web and downloadable PDF versions of the complete hybrid protocol, and a one-page abridged PDF version of the hybrid protocol. Videos allowed us to demonstrate procedures that were more difficult to communicate in text, such as how to determine the sex of the tree, how to upload data either in the iNaturalist app or on the iNaturalist website, and how to construct a cardboard sandwich.

Developing the user interface

While all virtual data were collected in iNaturalist, the project details and protocol materials were presented via a website we hosted (Smithsonian, n.d.). Screenshots of the Fossil Atmospheres leaf survey website page are included in the supplemental material (Appendix D). An embedded introductory video—a call to action that covered the broad project context—led the leaf-survey landing page. Below this video, text presented an overview of the project and outlined the materials needed to participate. The remainder of the page provided multiple methods for accessing step-by-step instructions for the hybrid protocol, including PDFs, a video, and the protocol presented in collapsible tabs. Introducing a web-based tabbed protocol not only provided mobile-friendly optimization for accessing instructions but gave users the flexibility to focus on the information they required. This made our protocol easier to navigate while allowing our audience to preferentially access information pertinent to their needs (Shneiderman et al. 2017). For example, some users might find detailed data entry instructions essential while others might find them unnecessary.

Once the protocol and website were developed, we engaged in multiple rounds of user testing with diverse groups of users for both the protocol and the website, engaging volunteers across age groups and with varying levels of comfort with technology. Their feedback led to several insights, including the need for a short, printable version of the protocol that participants could take outdoors with them, removing the need to switch screens between the digital protocol and the iNaturalist app.

We strove to meet standards of universal usability by accommodating alternate user scenarios. We expected most citizen scientists would participate through the iNaturalist mobile application but included instructions on how to submit observational data through iNaturalist's website to accommodate those without smartphones or those who might lack smartphone connectivity owing to limited data or infrastructure. Access to the internet was a requirement for participation. We did not provide a method of accepting data outside of the iNaturalist platform and we did not provide a method for receiving samples outside of the mail. We did have a few instances of people contacting the Smithsonian asking to deliver their samples in person and we did our best to accommodate these requests. To actively promote universal usability, we also included in the protocol an offer to mail people all the supplies and vouchers needed to return a sample to us via a package service free of charge. A handful of people requested vouchers. No one asked for mailing supplies.

Developing robust data practices

It was essential that we develop a robust plan for processing physical samples once they were received at the NMNH. Upon receipt we needed to protect and prepare the sample for acquisition while facilitating pairing the physical sample with its digital iNaturalist record. Upon arrival, a sample package was immediately marked with an internal project code. We then opened the package and extracted the cardboard sandwich. We entered the information noted on the package (iNaturalist username, date, and time of the observation) into an internal project database using the internal project number. We then opened the sandwich and tagged each leaf with that same internal project number. In this way, the physical samples were disassociated from personally identifiable information contained on the packaging. The leaves were then preserved for later scientific analysis.

Once the physical samples were secured within the repository of the NMNH Paleobiology collection and properly accessioned into the collections and museum database, the information provided by the participant on the cardboard sandwich had to be matched to a unique digital observation in the Fossil Atmospheres iNaturalist

project. If our hybrid protocol had been followed, this was straightforward. When participants departed from the protocol this became more challenging. We report on these challenges and our success at overcoming them in the Results section.

ENGAGE

Our promotion efforts began at the start of our August collection period and continued for approximately three weeks. We worked to identify potential partners that could help us communicate the project to broad audiences and engaged consistently across social media. We found particular success through Facebook, posting our call-to-action video on the platform on August 2nd and then re-engaging through that video and other content every 5–7 days throughout the month. Fossil Atmospheres also had active social media accounts on Twitter and Instagram. One member of the research team was assigned to each social media account and worked to be highly engaged and responsive on their assigned platform throughout the month. We also had a designated email account for the project through which we responded to requests for clarification or other help from more than 90 people.

RESULTS

We received 562 *Ginkgo biloba* samples (a single sample contained 6 leaves from a single shoot within a single cardboard sandwich) from 351 participants representing 37 states and 6 countries (*Figure 4*). Most participants

(77%) submitted one sample – however the project also attracted group efforts from school classes and “BioBlitz”-type biodiversity surveys. Nearly all participants (93%) entered all their data in one day, even if they reported multiple observations. Five percent of participants added sample data to iNaturalist over two (not necessarily consecutive) days. Five participants, representing 1.5% of total participants, uploaded data across 3 days. Our most highly engaged participant, with 29 sample submissions, uploaded data on 6 separate days.

We received 89% of samples within the August collection period. Samples received after August were also accepted to allow for collections made during August that were not mailed or received until the following month. It took, on average, 10.3 days for samples to be received at the NMNH after data for the sample were entered into iNaturalist (median: 8 days, range: 2–71 days). The high end of this range reflects the delays required by additional safety protocols for international samples. Samples were received throughout the project, but the highest weekly proportion was in the third week of the collection period (156 samples, or 28%).

PAIRING PHYSICAL SAMPLES WITH THEIR ONLINE RECORD

We received 562 physical samples and 608 digital observations. Multiple observations were made in the iNaturalist Fossil Atmospheres project for which no physical paired sample was received. We also received physical samples that did not have an associated digital record. Of the 562 physical samples received, 367 (60%) were easily

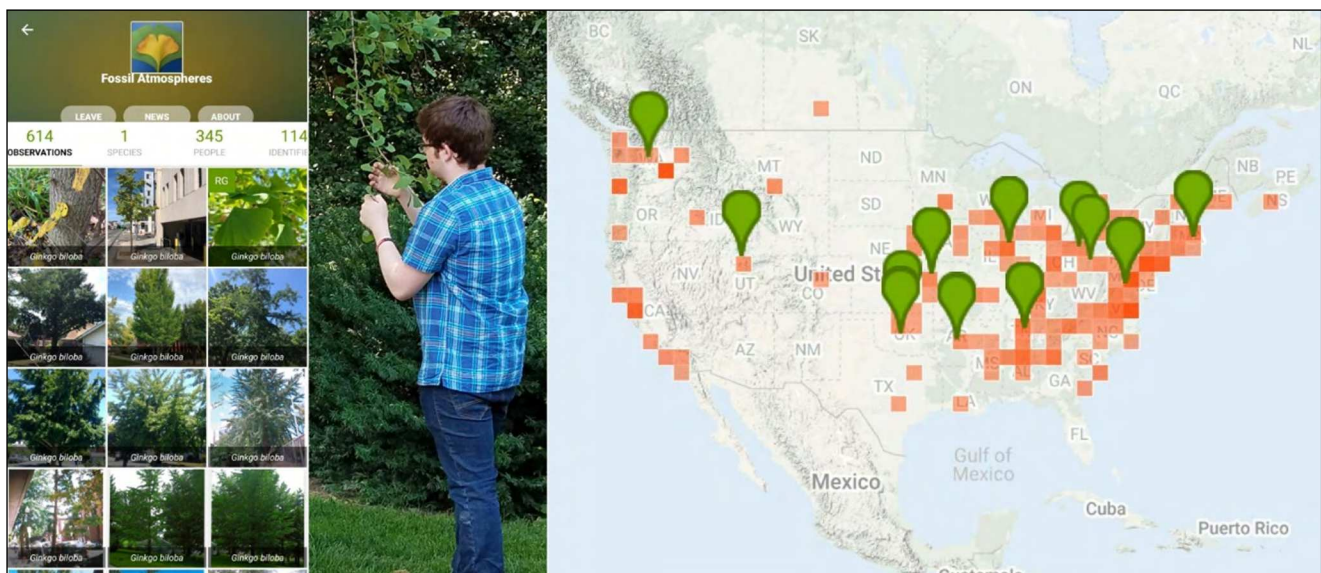


Figure 4 iNaturalist website user experience. a) iNaturalist app screenshot with scoreboard and recently submitted images. b) Citizen scientist collecting leaves. c) Distribution map on iNaturalist desktop homepage for Fossil Atmospheres; recent submissions highlighted with pins; older submissions generalized to regional squares.

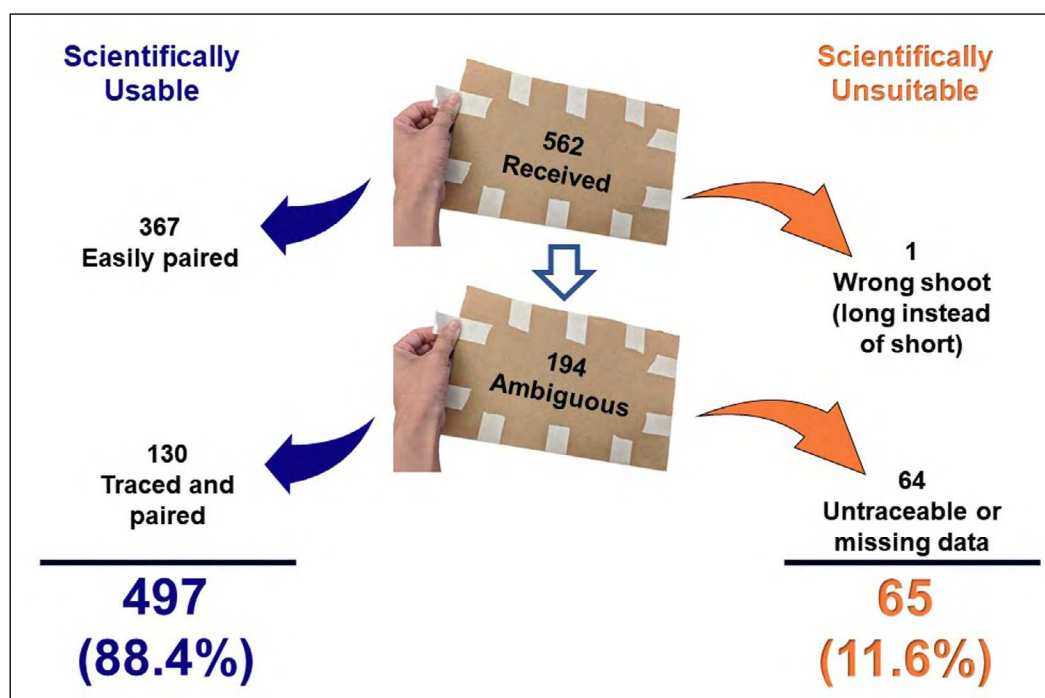


Figure 5 Summary of physical samples received by the Fossil Atmospheres project. Scientifically usable samples matched strict criteria, and totaled 88% of the 562 samples received.

paired with project-linked observations on iNaturalist. This left 195 physical samples and 241 digital records that were ambiguous. After additional investigation by the research team 65 samples (12%) were found to be unpairable or paired but missing key scientific data. We concluded with 496 samples (88% of all physical samples received) that met our scientific requirements and were able to be included in the Fossil Atmospheres sample set (Figure 5).

DATA INTEGRITY

The leaves of ginkgo have a unique and highly recognizable morphology, making it easy for both citizen scientists and iNaturalist's computer vision algorithms to correctly identify them. Our project benefited from this; 100% of the digital and physical samples we received were correctly identified as ginkgo. Only one sample, composed of leaves from the wrong part of the branch (see Appendix B), did not meet project requirements. Participants were very successful at following the protocol requirements for packaging and sending the physical sample. Of the physical samples received, 558 (99%) had complete or traceable information included on the cardboard sandwich (Table 1). In the instances where the information was incomplete or different from the iNaturalist observation, the name, return address, or postmark from the outer packaging could often be used to facilitate pairing the physical sample to an online observation. If more clarification was needed, we messaged users directly through the iNaturalist platform.

Using an established online platform as a data collection

tool gave a high level of confidence for digital data. Drop-down data fields requesting the height and sex of the tree were required before the iNaturalist observation could be submitted, and so were uniformly complete. Location and time information were automatically generated by the platform and so these fields were also uniformly complete.

Participants were more challenged by digital data entry than by the leaf shipping procedure. Many participants created multiple digital observations for the same physical sample: 46 records of the 241 ambiguous iNaturalist records were in effect duplications that could be collapsed into one observation and then included in the project. A more serious, but possibly correctable, challenge with the digital data involved citizen scientists entering their tree data on iNaturalist without linking that observation to the Fossil Atmospheres project. Because a standard iNaturalist observation requires only species identification, photo, location, and time of observation, entries that were not linked to our project did not contain responses to our required project fields. It was often possible to find these unlinked observations through a direct search of iNaturalist using either the project-specific data they included on their cardboard sandwich packaging or by using information from their outer mail packaging. As administrators of the iNaturalist project, the research team was then able to link these observations to the project digitally. However, that did not resolve the missing data. We could subsequently reach out to specific participants using iNaturalist's direct messaging feature to prompt them to provide the

| | NUMBER OF RECORDS RECEIVED | PERFECT RECORD | ONE OR MORE PROBLEMS | | FINAL RESULTS | |
|-------------------|-------------------------------------|-------------------|--|--------------------|---------------------|--|
| | | | UNUSABLE (MISSING DATA OR DUPLICATE) | USABLE (TRACED) | ACCEPTED SAMPLES | SCIENTIFICALLY USABLE SAMPLES (MET STRICT REQUIREMENTS) |
| iNaturalist entry | 608 | 367 | 111 | 130 | 497 (88%) | 496* (88%) |
| Physical sample | 562 | 554 | 4 | 4 | 558 (99%) | |

Table 1 Summary of records related to iNaturalist and sample return. *One sample contained leaves from an incorrect location on the tree and was rejected.

missing tree data. These additional efforts allowed us to successfully pair another 130 physical samples to their associated digital data, increasing the usable sample set from 367, which were perfect upon arrival, to 497 that were finally accepted by the project (**Table 1**).

On the rare occasions when participants mailed us a sample but did not generate any iNaturalist observation, there was no option for correction. In these cases, we knew the participant had read and followed the protocol, as evidenced by a correctly collected and packaged sample. Occasionally some or all of the required digital data were written on the cardboard sandwich, but more often, these samples presented without any paired tree data and we were therefore not able to include them in the data set.

DISCUSSION

The scientific aim for our citizen science effort within Fossil Atmospheres was to generate a large collection of high-quality ginkgo leaf samples with robust associated site data. Though all sampled localities were valuable contributions to our research, our primary goal was to cover, at minimum, three important north-south climatic transects across the contiguous United States. We also wanted to engage citizen scientists across the country with the project. As we began our collection month, the research team agreed that if we could get minimum coverage over our predetermined geographic transects and 100 users engaged with the project, we would consider the citizen science component of Fossil Atmospheres a success. We ended up achieving dense coverage over our geographic transects and engaged 345 participants in 37 states and 6 countries, who submitted 562 ginkgo cardboard sandwiches. This high engagement resulted in 88% scientifically usable samples, of total physical samples received. Our citizen science effort created a rich variety of leaf samples to support our investigation of climate change through the proxy of stomatal index. Our results also compare favorably to other national projects that collect physical samples, such as the Harvard Personal Genome Project, which reports that of

1,143 users, 185 produced complete genomes, equating to a 16.2% rate of return (Ball et al. 2014).

In addition to our scientific goal, we wanted to explore the benefits of using an established online platform with well-developed functionality—including both mobile and web interfaces and a large, pre-existing community of contributors—in a carefully designed hybrid collection protocol in which participants submitted photos and site data digitally and physical samples using the mail. The project team experienced no technical issues while using iNaturalist as a data collection tool, highlighting the value of working with an established platform that employs their own skilled technology team. This freed us from technical troubleshooting or oversight efforts. We found the team at iNaturalist available and responsive via email whenever we had questions. We also made use of the extensive iNaturalist support documentation available online.

DESIGNING A HYBRID PROTOCOL

We realized success at all three phases of project development and implementation (**Figure 2**). We present these actions as a template that other projects may follow.

Choose

Selecting a sampling period is project specific. A four-week period was sufficient for us. Similarly, choosing an online platform best aligned with set scientific goals is also project specific. How project data is stored and accessed through a pre-existing platform should be considered carefully. As an example, once we were prepared to download our complete data records from iNaturalist we were confronted with an unanticipated challenge regarding the image files. The iNaturalist platform was designed to provide easy access to non-image data. However, standard data downloads provided hyperlinks to the images, which did not meet our project's requirement to house all our data in NMNH systems.

Design

Protocols should be as streamlined as possible with single steps that ideally collect multiple types of data. We

recommend that instructional guides are presented in a variety of formats, including explanatory videos, printed text within a PDF, and formatted text on a website with text guides in both complete and abbreviated versions. We found a central, dedicated webpage effective. A central website allowed us to share engaging visuals to present a strong and consistent brand, print resources that could be used to promote our project, and the hybrid protocol in various forms that attended to universal usability.

Robust data practices are also essential. For a hybrid protocol, we found this to mean designing a data management plan for accessing and storing project data, a protocol for receiving physical samples, and a protocol for linking the physical samples to their corresponding digital data. All protocols must be designed to protect the privacy of participants.

Engage

Engagement becomes essential when the project is ready to be launched. We found being active and responsive on social media, along with answering specific questions via a dedicated project email, was a successful strategy.

LIMITATIONS OF ONLINE PLATFORMS

Online platforms are extremely useful for organizing the collection of data but have profound limitations for long-term data storage. Although it would be tempting to consider a professionally maintained network like iNaturalist as a permanent repository for data, such a practice would not be considered good data management. Internet-based databases may have a lifespan and are accessible to outside users for only as long as the sites are functional online. Databases cannot be guaranteed to be backed up, engulfed, or accessible in perpetuity. Online platforms for long-term data storage also frequently lack database versioning that allows previous database iterations to be referenced later. Though download dates and version numbers are standard to report, there is no guarantee a reported dataset would be reproducible, as iNaturalist and other online databases are constantly under development and growing.

Lastly, how and what data is accessible is dictated by the online platform, which may lead to unanticipated restrictions. For example, iNaturalist limits file download size and suggests that users utilize the Global Biodiversity Information Facility (GBIF 2021) database for larger downloads. This is reasonable, but there are caveats. iNaturalist uploads only “research-grade” observations to the GBIF website. Our observations of *Ginkgo biloba* are considered by iNaturalist to be a “casual” observation, because effectively all specimens of current ginkgo were planted. Despite the fact that Fossil Atmospheres is

collecting these casual observations for research purposes and that we have acquisitioned the physical specimens into our collections at the NMNH, *Ginkgo biloba* records are not included in databases like GBIF, which exists to document natural diversity patterns. For these reasons, we highly recommend that researchers create their own permanent and versioned copy of data when utilizing online platforms for citizen science, as we have for these collections, which are permanently housed at the NMNH.

LIMITATIONS OF THE PROTOCOL

A primary limitation of our hybrid protocol involved universal usability. We took care to design alternative paths for participation that did not require a smartphone, but at minimum, participation did require a camera, a computer, and internet access. We had 24 participants send us physical samples that were not associated with iNaturalist in any way. Some participants wrote some, or even all, of the required digital data on the cardboard sandwich, but without photos and a confirmed location, these samples could not be included in the Fossil Atmospheres data set. We can’t know why these participants chose to eschew the online platform tool. At this time, it is also not possible to know how many people were discouraged from participating by the technological requirements.

Our hybrid protocol also did not allow us to collect information on who the citizen scientists were or exactly how they interacted with the protocol. For example, we were unable to determine which form of the protocol guide participants found most useful or if they were using the app or the website to enter their digital data. This information could be collected by future projects by including unique data fields that asked demographic questions or queried the user experience. Further analysis, which is outside the scope of this paper, could also give some indication of whether participation in the project impacted people’s use of the online platform—for example, how many participants were regular or first-time users of iNaturalist and whether, when a ginkgo observation was their first action on the platform, they stayed and became regular users.

CONCLUSIONS

Our hybrid citizen science protocol worked to combine physical samples and digital observations. Here we report on one part of the Fossil Atmospheres project—a project that uses ginkgo leaves to investigate climate change—and share our strategy when designing, and then successfully implementing, a citizen science effort using a hybrid protocol. Our project is not alone in pairing physical sample collection with digital data, e.g., the fungal diversity project

FunDiS (Sheehan et al. 2021); however, how this success is achieved is rarely reported in detail. The hybrid protocol we detail here supported participants in collecting photos and in giving accurate species and location data as well as additional site data; as a result, 88% of all physical samples received met our project's scientific requirements.

We further found that using an established online platform as a technological tool for data collection had multiple benefits. Unique data fields were simple to set up, and data input was user-friendly thanks to the professional user experience supplied by the established platform. Metadata, such as date, time, and location of an observation, were collected automatically, while computer vision coupled with expert community crowdsourcing verified species identity. This resulted in 100% of the received samples being ginkgo.

We recognize that online platforms are constantly evolving and that the details of our experience may quickly be out of date. We identified from our experience, however, a three-phase project model that can provide a clear path toward a successful large-scale hybrid project that collects high-quality citizen science data. Our project also provides a further test case (one that includes citizen scientists) of a general model (Heberling and Issac 2018) that is broadly applicable to projects wishing to enhance the data associated with museum specimens.

DATA ACCESSIBILITY STATEMENT

Citizen science data is provided in a supplementary file to this paper (Supplemental File 5). The Fossil Atmospheres Project remains accessible on the iNaturalist platform. Researchers needing more information can contact the lead author to gain access to the collections and databases at NMNH.

SUPPLEMENTARY FILES

The supplementary files for this article can be found as follows:

- **Supplemental File 1: Appendix A.** Screenshots of the Fossil Atmospheres iNaturalist project page. DOI: <https://doi.org/10.5334/cstp.422.s1>
- **Supplemental File 2: Appendix B.** Our complete hybrid protocol. DOI: <https://doi.org/10.5334/cstp.422.s2>
- **Supplemental File 3: Appendix C.** Screenshot of the Fossil Atmospheres leaf survey website page. DOI: <https://doi.org/10.5334/cstp.422.s3>
- **Supplemental File 4: Appendix D.** Screenshots of the Fossil Atmospheres leaf survey website page. DOI: <https://doi.org/10.5334/cstp.422.s4>

- **Supplemental File 5.** Citizen science data. DOI: <https://doi.org/10.5334/cstp.422.s5>

ETHICS AND CONSENT

Our citizen science project was reviewed by the Smithsonian Institution's Institutional Review Board under NIST 800-53.

ACKNOWLEDGEMENTS

Our success designing and implementing this hybrid citizen science protocol was only possible through the help of many scientists, educators, Smithsonian volunteers, and communicators. We would like to particularly acknowledge: Robert Costello for facilitating institutional support; Kathy Hollis and Scott Wing for accessioning leaves into the Paleobiology collections; Tina Tennesen and her team at NMNH for social media support; Sophia Roberts and her team for video production support; Sue Lutz for permission to film in the NMNH herbarium and help with international sample acquisition; Merijke Coenraad and Lautaro Cabrera for user testing; Grace Biggs, Pam Hamilton, and Luke Gimmelbein for sample processing and data validation; Carlita Sanford, Skip Lyles, and the NMNH mail room team for help with receiving physical samples; and, most importantly, the hundreds of wonderful citizen scientists that collected samples for our project.

FUNDING INFORMATION

This work was funded by The National Science Foundation under EAR:1805228.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

RSB, LCS, LMC, and HK conceived the project, designed the protocol and website, designed the instructional materials and data practices, and implemented the sampling period. HK and RSB conducted the analysis of the sample data. HK drafted the manuscript. All authors contributed to the final version of the paper.

AUTHOR AFFILIATIONS

Heather Killen  orcid.org/0000-0001-6922-8116

University of Maryland, College Park, US

Lucy Chang  orcid.org/0000-0001-9316-5290

Royal Ontario Museum, CA

Laura Soul  orcid.org/0000-0001-5352-886X

Natural History Museum, London, GB

Richard Barclay  orcid.org/0000-0003-4979-6970

Smithsonian National Museum of Natural History, US

REFERENCES

- Aristeidou, M, Scanlon, E and Sharples, M.** 2017. Profiles of engagement in online communities of citizen science participation. *Computers in Human Behavior*, 74: 246–256. DOI: <https://doi.org/10.1016/j.chb.2017.04.044>
- Ball, MP, Bobe, JR, Chou, MF, Clegg, T, Estep, PW, Lunshof, JE, Vandewege, W, Zaranek, AW and Church, GM.** 2014. Harvard Personal Genome Project: lessons from participatory public research. *Genome medicine*, 6(2): 1–7. DOI: <https://doi.org/10.1186/gm527>
- Barclay, RS and Wing, SL.** 2016. Improving the Ginkgo CO₂ barometer: implications for the early Cenozoic atmosphere. *Earth and Planetary Science Letters*, 439: 158–171. DOI: <https://doi.org/10.1016/j.epsl.2016.01.012>
- Bonney, R, Phillips, TB, Ballard, HL and Enck, JW.** 2016. Can citizen science enhance public understanding of science?. *Public Understanding of Science*, 25(1): 2–16. DOI: <https://doi.org/10.1177/0963662515607406>
- GBIF: The Global Biodiversity Information Facility.** (2021). What is GBIF?. Available from DOI: <https://www.gbif.org/what-is-gbif> [31 March 2021].
- Groom, Q, Strubbe, D, Adriaens, T, Davis, A, Desmet, P, Oldoni, D, Reyserhove, L, Roy, HE and Vanderhoeven, S.** 2019. Empowering citizens to inform decision-making as a way forward to support invasive alien species policy. *Citizen Science: Theory and Practice*, 4(1): 1–11. DOI: <https://doi.org/10.5334/cstp.238>
- Heberling, JM and Isaac, B.** 2018. iNaturalist as a tool to expand the research value of museum specimens. *Applications in Plant Sciences*, 6(11): e01193. DOI: <https://doi.org/10.1002/aps3.1193>
- iNaturalist.** n.d. iNaturalist Observations, n.d. Available at <https://www.inaturalist.org/observations> [Last accessed 9 February 2022].
- Kosmala, M, Wiggins, A, Swanson, A and Simmons, B.** 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14(10): 551–560. DOI: <https://doi.org/10.1002/fee.1436>
- La Sorte, FA and Somveille, M.** 2020. Survey completeness of a global citizen-science database of bird occurrence. *Ecography*, 43(1): 34–43. DOI: <https://doi.org/10.1111/ecog.04632>
- MacPhail, VJ and Colla, SR.** 2020. Power of the people: A review of citizen science programs for conservation. *Biological Conservation*, 249: 108739. DOI: <https://doi.org/10.1007/s10841-019-00152-y>
- myFOSSIL.** (n.d.) Social Paleontology, n.d. Available at <https://www.myfossil.org/> [Last accessed 20 November, 2021].
- Newman, G, Wiggins, A, Crall, A, Graham, E, Newman, S and Crowston, K.** 2012. The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, 10(6): 298–304. DOI: <https://doi.org/10.1890/110294>
- Pandya, RE.** 2012. A framework for engaging diverse communities in citizen science in the US. *Frontiers in Ecology and the Environment*, 10(6): 314–317. DOI: <https://doi.org/10.1890/120007>
- Parker, SS, Pauly, GB, Moore, J, Fraga, NS, Knapp, JJ, Principe, Z, Brown, BV, Randall, JM, Cohen, BS and Wake, TA.** 2018. Adapting the bioblitz to meet conservation needs. *Conservation biology*, 32(5): 1007–1019. DOI: <https://doi.org/10.1111/cobi.13103>
- Peterman, K, Bevc, C and Kermish-Allen, R.** 2019. Turning the King Tide: Understanding dialogue and principal drivers in an online co-created investigation. *Citizen Science: Theory and Practice*, 4(1). DOI: <https://doi.org/10.5334/cstp.189>
- Seltzer, C.** 2021. What is it?, 8 February, 2021. Available at <https://www.inaturalist.org/pages/what+is+it> [Last accessed 28 November 2021].
- Sheehan, B, Stevenson, R and Schwartz, J.** 2021. Crowdsourcing Fungal Biodiversity: Approaches and standards used by an all-volunteer community science project. *Biodiversity Information Science and Standards*, 5: e74225. DOI: <https://doi.org/10.3897/biss.5.74225>
- Shneiderman, B, Plaisant, C, Cohen, MS, Jacobs, SM and Elmqvist, N.** 2017. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6th. ed.). Hoboken, NJ: Pearson Education, Inc.
- Simpson, R, Page, KR and De Roure, D.** 2014. Zooniverse: observing the world's largest citizen science platform. *Proceedings of the 23rd International Conference on World Wide Web*, 1049–1054. DOI: <https://doi.org/10.1145/2567948.2579215>
- Smithsonian.** n.d. Citizen science: Leaf survey—A component of the Fossil Atmospheres project, n.d. Available at <https://www.si.edu/fossil-atmospheres/leaf-survey> [Last accessed 19 March 2021].
- Soul, LC, Barclay, RS, Bolton, A and Wing, SL.** 2019. Fossil Atmospheres: a case study of citizen science in question-driven palaeontological research. *Philosophical Transactions of the Royal Society B*, 374(1763): 20170388. DOI: <https://doi.org/10.1098/rstb.2017.0388>

- Spellman, KV and Mulder, CP.** 2016. Validating herbarium-based phenology models using citizen-science data. *BioScience*, 66(10): 897–906. DOI: <https://doi.org/10.1093/biosci/biw116>
- Sullivan, BL, Aycrigg, JL, Barry, JH, Bonney, RE, Bruns, N, Cooper, CB, Damoulas, T, Dhondt, AA, Dietterich, T, Farnsworth, A and Fink, D.** 2014. The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation*, 169: 31–40. DOI: <https://doi.org/10.1016/j.biocon.2013.11.003>
- Unger, S, Rollins, M, Tietz, A and Dumais, H.** 2020. iNaturalist as an engaging tool for identifying organisms in outdoor activities. *Journal of Biological Education*, 1–11. DOI: <https://doi.org/10.1080/00219266.2020.1739114>
- US General Services Administration.** n.d. Step 4 – Managing Your Data, n.d. Available at <https://www.citizenscience.gov/toolkit/howto/step4/#> [Last accessed 20 February 2021].
- Wittmann, J, Girman, D and Crocker, D.** 2019. Using iNaturalist in a coverboard protocol to measure data quality: Suggestions for project design. *Citizen Science: Theory and Practice*, 4(1). DOI: <https://doi.org/10.5334/cstp.131>
- Woodward, FI.** 1987. Stomatal numbers are sensitive to increases in CO₂ from pre-industrial levels. *Nature*, 327(6123): 617–618. DOI: <https://doi.org/10.1038/327617a0>

TO CITE THIS ARTICLE:

Killen, H, Chang, L, Soul, L and Barclay, R. 2022. Combining Physical and Digital Data Collection for Citizen Science Climate Research. *Citizen Science: Theory and Practice*, 7(1): 10, pp. 1–14. DOI: <https://doi.org/10.5334/cstp.422>

Submitted: 31 March 2021 Accepted: 10 January 2022 Published: 09 March 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Citizen Science: Theory and Practice is a peer-reviewed open access journal published by Ubiquity Press.