A peer production model for citizen science: comparative analysis of three online platforms

Katharina Kloppenborg^{1,4,*}, Mad Price Ball^{2,*}, and Bastian Greshake Tzovaras^{1,2,3*}

¹Learning Planet Institute / Inserm U1284, Université Paris Cité, Paris, France

²Open Humans Foundation, United States

³The Alan Turing Institute, London, United Kingdom

⁴Lead contact

*Correspondence: katharina.kloppenborg@cri-paris.org (K.K.), mad@openhumans.org (M.P.B.), bgreshaketzovaras@turing.ac.uk (B.G.T.)

Summary

As a data-based practice, citizen science involves volunteers in various stages of research, often including the collection and processing of data through online platforms. Citizen science projects frequently combine epistemic goals with societal benefit goals, such as democratizing science. However, it is not well understood how to design such participation architectures to enable volunteer engagement in support of these goals. In this study, we introduce a working model of peer production for citizen science for analysis and comparison of key design features for participation and their consequences. To test this model, we apply it to three citizen science case studies and show that it helps move beyond existing classification approaches, supports the structured analysis of key features across different approaches, and their implications. This work represents a step towards understanding and informed decision-making of participation architectures for existing as well as future online infrastructures for citizen science.

Data Science Maturity

DSML 1: Concept

Keywords

citizen science, peer production, participation architecture, data infrastructures, social computing

Introduction

Citizen science is a growing practice across various academic disciplines, with a frequent focus on aggregating and processing large datasets. As such, citizen science is a field that employs data science techniques to guide the collection and analysis of data. Beyond its use as a format for data collection, citizen science is also seen as having potential societal benefits, including through improving scientific literacy¹, democratizing science², improving data justice³, and enabling “undone science”, which refers to areas of research that are considered important by civil society, but that lack funding⁴. As a result, citizen science projects often combine epistemic objectives with societal benefit goals, and can differ considerably in how they balance and prioritize these.

The ways in which volunteers participate in citizen science projects also vary widely – indeed, the term “citizen science” itself has been used to describe very different contexts, and there is often a lack of consensus regarding when the term is appropriate⁵. Among projects labeled as citizen science, the most common form are “contributory” projects⁶, alternatively described as “crowdsourcing”⁷, where large numbers of volunteers contribute to data collection or processing in projects designed by professional scientists. While the scope of volunteer involvement is limited in such projects, other forms of citizen science include volunteers in other aspects of research, like problem definition or data interpretation. Bottom-up citizen science projects may be initiated and accompanied by institutions or professional researchers in all or most steps, like in the case of “Extreme Citizen Science” initiatives, which collaborate for example with indigenous communities in conservation projects⁸. Projects may also be initiated as community-led “grassroots” activities, with varying degrees of institutional involvement in project design and management, for example in environmental justice⁹, or patient advocacy¹⁰. This diversity in types of participation implies diversity in how projects are designed to engage participants.

Many citizen science projects rely heavily on online platforms and tools for collaborative knowledge production, and these platforms are themselves a subject of sociological and design research. Participation architectures, i.e. the opportunity structures available to external contributors¹¹, can vary greatly depending on the projects' goals and participation concepts. Such meta-research efforts on citizen science platforms have mostly focused on crowdsourcing aspects to investigate user behaviors, for example as they relate to learning norms of practice¹², coordinated action¹³, and data quality¹⁴, or to derive design claims¹⁵ and guidelines¹⁶ for creating successful projects. While there are dedicated platforms that support community-led citizen science projects, such as Public Lab¹⁷, they typically operate at a smaller scale than crowdsourcing platforms, and to date there has been less meta-research about them. In addition, bottom-up initiatives usually operate on a smaller scale, often developing or adapting data collection and analysis tools to their local context⁸.

Interestingly, among the works focusing on crowdsourcing, several suggest that projects may benefit from deeper participant engagement in pursuit of epistemic goals: They suggest providing features that help newcomers move from peripheral tasks to more impactful work¹², and new infrastructures that support more collaborative and advanced work, authority and legitimacy¹³. They argue with the proven competence of citizen scientists to handle complex tasks, like identifying glitches while image tagging, or resolving conflicts when determining new taxonomies¹³, and cite scientific discoveries, e.g. in astronomy, that have been achieved by citizen scientists engaging beyond the requirements of the predefined task^{13,15}. They go even further, suggesting redefining the role of volunteers as scientific assistants¹³, or establishing communities that function semi-autonomously¹⁵. Additionally, there have been calls for deeper, more meaningful engagement, referring to the potential to transform what knowledge is created, how it is created, and by whom². Regardless of motivation, the question of how to create participation architectures that enable greater volunteer engagement remains largely unanswered in literature. This poses significant barriers to the design or redesign of existing and future citizen science platforms. Practitioners need tools and guidance to help them gain a systematic understanding to make informed design decisions that are best suited for their topic and project goals.

To address this gap and provide practitioners with the needed support, we here introduce a model that operationalizes and applies the characteristics of “peer production” theory to the larger citizen science context.

Peer production as a mode of distributed knowledge production relies on self-organizing communities of individuals who collaborate online¹⁸ and is widely known from its applications in Wikipedia and open-source software development. Knowledge creation in citizen science often shows similarities to peer production: Volunteers with different skills and backgrounds produce knowledge artifacts collaboratively, typically online as a distributed community of individuals and with non-monetary motivations.

Given these similarities, there have been calls for peer-produced citizen science¹⁹ and we agree that peer production theory has the potential to help facilitate multifaceted engagement of volunteers in citizen science: In particular, peer production approaches enable new contributors to progress from simple routine tasks to complex, creative work, and to switch back and forth between them at their free choice, e.g. from micro tasks like fixing typos to complex tasks like writing software²⁰, reflecting the call for citizen science features to move newcomers from peripheral tasks to more impactful work^{12,13}.

As most peer-production literature related to citizen science tends to focus on political aspects, governance or intellectual property^{21,22}, we build our working model of peer production as a summary of key characteristics of online platforms that enable the peer production of knowledge, applied to the citizen science domain.

To demonstrate the use of the model, we apply it to three case studies – the citizen science platforms Zooniverse, iNaturalist, and Public Lab. We characterize and compare the platforms' participation architectures, and discuss the impact of different design choices.

In doing so, we intend to give citizen science practitioners a tool at hand to reflect participation architectures of existing and future platforms, and to provide support for taking informed design decisions with respect to peer production characteristics.

Results

Working Model of Peer Production

This working model is an effort to operationalize peer production for citizen science. It consists of a set of key characteristics of collaborative online knowledge production. It is intended to be used to analyze citizen science online platforms that offer functionality for user contributions.

The model was derived from existing peer production literature, largely from Kostakist and Bauwens²³, and Benkler²⁰. More details, and information on the background of this model can be found in the section *Experimental procedures* and *Supplementary Material A*. Figure 1 summarizes and illustrates the working model of peer production, including examples from Wikipedia, as one of the most successful and widely known peer production platforms. Links to the sources of the examples can be found in *Supplementary Material B*.

Peer production characteristic		Wikipedia example
Common research object		Open encyclopedia
Range of tasks		Creating, editing, discussing articles
Granularity and modularity		Articles, sections, sentences, ...
Equipotential self-selection		Users choose what and how much they want to edit
Quality control		Talk pages for feedback, rollback function to undo changes
Learning trajectories		Manuals, suggestions of easy tasks for beginners
Direct and indirect coordination		Flags like "citation needed", red links, talk pages

Figure 1. Working model of peer production with examples from Wikipedia

A central feature of the model is the **common research object** that is collectively produced. It results from the goals of the project. A project can have one or more common research objects. Since they are something that is created collaboratively, they are usually composed of many parts (see “granularity and modularity”). In Wikipedia, the common research object is an open encyclopedia.

The **range of tasks** describes what types of work are accessible to users, i.e. tasks of which quality, size, and complexity. For online platforms, this can be analyzed by inspecting the interaction possibilities of the user interfaces. In Wikipedia, the range of tasks includes, amongst many more, creating, editing, or discussing articles.

Granularity and modularity refers to how tasks and the research object are broken down into smaller tasks and objects. Related is the cost of integration, which describes the effort required to integrate small elements into larger modules that make up the research object. On Wikipedia, the encyclopedia is made up of modules called articles, which in turn are composed of sections, sentences, etc. The editor allows low-effort integration of edits or new pages into the system. Creating a consistent narrative or putting articles in the right categories can equal higher integration effort.

Equipotential self-selection relates to which tasks are accessible to regular platform users, and to what extent they can choose and self-assign tasks. On Wikipedia, users can choose to create or edit any article of their choice, without the need to prove expertise with formal entry credentials, like a degree in a relevant field. They can also decide how much and which parts of an article they would like to edit, without waiting for anyone to assign them what to work on.

The **quality control** characteristic covers the mechanisms in place to judge the quality of user contributions. If done by the user community, it can also be referred to as “communal validation”²³. On Wikipedia, for example, users can give feedback on the “Talk” page of an article. There is also a rollback function to go back to former states, and users can become stewards for specific pages and topics and watch in more detail over their development.

Learning trajectories refers to how a newcomer learns to execute tasks, and which possibilities they have to learn and perform more advanced tasks. On Wikipedia, there are many guides to teach newcomers, for example the “Manual of Style”, and to document consensus on design guidelines. New users have the possibility to pick micro-tasks, like editing

spelling errors, before starting bigger tasks, and can thus learn by practice. There are also community guidelines in favor of treating newcomers or other users who might make mistakes with patience and benevolence (“Assume good faith”).

Direct and indirect coordination relates to features that help coordinate work, e.g. that signal open tasks, as well as features that allow for direct communication between users, like chat systems. In Wikipedia, for direct coordination, there are talk pages for articles and user pages. Features to signal open tasks are, for example, links in red color that indicate that a page that is linked in another page has not been created yet, section headings followed by little or no text, or flags like “citation needed”, or “stub”, which can be integrated into articles.

Case Studies

To evaluate the working model of peer production, we apply it to three citizen science platforms to identify how the different characteristics are implemented: Zooniverse, iNaturalist, and Public Lab. All three of these are online citizen science platforms with active user bases and jointly cover a wide range of different participation approaches that can be found in citizen science. A schematic representation that describes the knowledge production processes for each platform can be found in Figure 2. Links to sources for the case studies can be found in *Supplementary Material B*.

Figure 2. Schematic representation of knowledge production in the case studies.

Zooniverse²⁴ is a popular platform for crowdsourcing citizen science. Volunteers assist professional researchers from all kinds of domains, by collectively classifying data for their projects. Zooniverse developed from the Galaxy Zoo project and is now a collaboration between the University of Oxford, the University of Minnesota - Twin Cities (UMN) and Chicago's Adler Planetarium. With more than 2.5 million registered users, and almost 700 million classifications as of September 2022, it is one of the biggest online platforms for citizen science.

iNaturalist²⁵ is a social network and online platform for sharing, mapping and identifying biodiversity observations around the world. While collaboratively building a biodiversity database, they frame their primary aim as “connect[ing] people to nature”. iNaturalist began in 2008 as a Master's final project at the UC Berkeley's School of Information and is now a joint initiative between the California Academy of Sciences and the National Geographic Society, with almost 5.7 million registered users and over 115 million uploaded observations.

Public Lab²⁶, short for Public Laboratory for Open Technology and Science, is a community, non-profit organization and online platform, aiming to empower people to address environmental justice issues through community science and open technology. They focus on the collaborative development, sharing, reuse and adaptation of DIY scientific tools intended to enable community investigation of environmental concerns. Public Lab was formed after the BP oil disaster at the Gulf of Mexico in 2010, when residents documented damage in a grassroots mapping effort using weather balloons¹⁷. There are over 2,100 registered users, who contributed numerous topic pages, reports, and manuals for do-it-yourself environmental monitoring projects.

Zooniverse

Common research object For nearly all projects on Zooniverse, the main research objects are large datasets of images, videos or text that need to be annotated and which are typically provided by professional researchers. Projects span a wide range of topics and users are asked to annotate individual data points such as telescope images of galaxies, create transcriptions of historic diary pages, or identify noises of deep sea animals from acoustic samples.

Range of tasks The main task type is the annotation of a data point according to the instructions given for each individual project. Beyond this, users can make collections of data points they classified, and the platform offers forums for each project as well as talk pages for each individual data point (see Figure 2A). These talk pages can be used for leaving comments,

hashtags, adding data points to collections, and for participating in discussions for quality control and learning (see below). Committed users can become moderators of talk forums for specific projects.

Granularity and modularity Depending on the project, annotation tasks can differ in complexity and duration, e.g. ranging from selecting from predefined options to transcribing documents. Within one project, the granularity of annotation tasks is of similar size. Modularity is given by the automatic integration of user contributions to the database after annotation.

Equipotential self-selection Users choose which projects they want to contribute to based on their interests, and the quantity of data points they want to process, within the limits of available data. Beyond this, users do not self-select tasks as the data points are automatically assigned to the user one at a time. To engage further, users can take part in discussions on projects or individual data points using the aforementioned talk pages. Users can become talk forum moderators after being granted the rights by the project leads. Creating new projects requesting annotation is usually reserved to professional scientists.

Quality control Annotations of other users are hidden and the main quality control of contributions is done by project leads behind the scenes without further input by the contributors. However, post-classification, the user can access the data point's talk page to provide feedback or discuss with other users who have annotated the same data point.

Learning trajectories For each project, tutorials are provided by project leads in order to teach users how to annotate for a given project. Further engagement is possible on talk pages. Users can take up more advanced roles, e.g. becoming a moderator in a project talk forum. These more advanced roles are assigned by the project leads.

Direct and indirect coordination There is no coordination between users in the main data processing workflow, because the annotations of other users are hidden. However, after annotating, users can access a page dedicated to the respective data point, where they can add tags and comments, and see meta-information other users entered. For further communication there is a forum attached to each project.

iNaturalist

Common research object The main research object of iNaturalist is a public biodiversity database consisting of sightings of species with metadata like location, time, and photos.

Range of tasks iNaturalist includes a similar range of tasks as Zooniverse, expanded by some functions (see Figure 2B). Beyond annotating data, which consists in adding species identifications to observations, additional tasks include the collection of new data points along with metadata, as well as the possibility of writing journal posts, discussing in a forum, and creating data collection projects that encourage users to provide observations specific to certain areas or species. Users in the curator role can edit the taxonomy, create and edit wiki pages, resolve flags on taxa and user content, and hide comments and suspend users for violating community guidelines.

Granularity and modularity In iNaturalist tasks are usually of small size, ranging from identifying existing observations to uploading new observations. Modularity is provided on two levels: Observations are automatically added to the database and provide the data pool for future annotations. Secondly, once the species has been collectively identified (see quality control), the observation is automatically added to the Global Biodiversity Information Facility (GBIF) database²⁷.

Equipotential self-selection Users can select from the range of tasks mentioned above. Users are free to decide the type and quantity of observations they would like to annotate or add to the database, without automatic assignment. There is a curator role that is granted by iNaturalist staff for users who have engaged continuously over several months: curators have additional rights, e.g. to edit iNaturalist's taxonomy or to hide comments that violate community guidelines.

Quality control There is a communal validation process for species identification, using the following mechanism: New observations are automatically tagged with "needs ID". As the observations and annotations are generally publicly accessible, users can see other users' annotations, and can agree, disagree, or make a guess more specific. For example, if the person uploading the observation specified it as "ladybird", other users can either confirm this identification, specify what species of ladybird it is or disagree by identifying it as a different insect. The resulting annotation is called "community taxon" (see screenshot in *Supplementary Material C1*). To achieve research grade, the ID needs to be supported by at least two users,

and either have a community taxon at species level or lower. This research grade database is publicly accessible under an open license but some metadata, like specific location, can be obscured, to protect the privacy of underage users, or to protect endangered species.

Learning trajectories To learn how to contribute to the database, there is a “Getting started” guide with tutorials on uploading observations and species identification. Users are free to address bigger tasks, like creating their own projects, or offline data-collection events (“bioblitz”) for which there are also help pages, or become curators after having collected experience through regular engagement.

Direct and indirect coordination There are features to coordinate work, like “needs ID” tag in the communal validation process, projects that call for specific observations, or filters to help users identify observations in areas of their expertise. For direct communication, there is a forum, as well as comment and blog functionality.

Public Lab

Common research object On Public Lab, the research object is a catalog of experimental designs and project reports in the domain of DIY environmental science. Unlike on Zooniverse and iNaturalist, there is no data collection or annotation mechanism on the platform. Instead, artifacts include collections about DIY environmental monitoring and knowledge transfer, like topic pages, similar to wiki pages, e.g. on air quality, and research notes, similar to blog posts (e.g. “How to Build a Bucket Air Monitor”). There are artifacts for problem solving, like issue briefs, in which users explain their issue to support from other users, or more general question elements that other users can reply to.

Range of tasks Reflecting the diversity of research artifacts, Public Lab users can work on a range of tasks that can be split roughly in two categories: looking for information on experimental design, and sharing information or outcomes (see Figure 2C). To look for information, users can create dedicated artifacts, i.e. questions and issue briefs, or consult other existing artifacts on their topic of interest. To share outcomes or other information, users can create or contribute to topic or project pages, research notes, DIY toolkits, and blog posts. They can replicate other users’ experiments (“activities”), comment and reply to questions, issue briefs, or research notes.

Granularity and modularity The wide range of tasks have varying granularity, ranging from small edits on topic wiki pages to uploading larger building blocks such as experimental designs. Modularity is achieved by interlinking contributions on Public Lab through tags and backlinks. These tags need to be added to an artifact to link it to a topic or artifact category, to add backlink functionality (“i did this” / “seeks:replications”, “activity:”, “replication:”, “activity:infragram”) to integrate their artifacts / make them findable. Tags are pre-specified if an artifact is created by clicking on “add an activity”, “I did this”, “ask a question” from a related page (see screenshot in *Supplementary Material C2*).

Equipotential self-selection Similar to iNaturalist, Public Lab users have access to all content to self-select open tasks they are most interested in or add new content to their liking. On application, users can take on moderator or admin roles.

Quality control Due to the free, textual form of the research artifacts, there is no algorithmic quality control that measures consensus like in iNaturalist or Zooniverse. Rather, feedback is generally in written form of comments to other artifacts, or in the form of new blog entries, when users replicate each other’s activities and report their protocol and how it worked in their case.

Learning trajectories For new users, there are so-called issue brief artifacts to remove barriers and encourage users to get started (see screenshot in *Supplementary Material C3 a*), “first-time-posters” tag, templates and forms for artifacts, and helpful instructions in default texts of text boxes (see screenshot in *Supplementary Material C3 b*). There are also question objects to help resolve specific issues. It is possible to replicate existing experiments or ask other users to replicate one’s own experiment to get feedback and learn from one’s own or other people’s experiences. More experienced users can apply for moderation or administration roles.

Direct and indirect coordination There are several features that help signal open work, like the “I did this” button (see screenshot in *Supplementary Material C2 a*), that appears on pages tagged with “activity”, which depicts an invitation for other users to replicate the experiment, and at the same time serves to connect the original and replicated activity. Other ways to signal that help is needed are dedicated artifacts like “questions” or “issue briefs”. For direct communication, there is a comment functionality under all types of artifacts.

Discussion

The goal of this study was to develop a systematic approach to understanding the characteristics of online platforms that enable deeper volunteer engagement in citizen science, following literature highlighting its potential for achieving epistemic or societal goals. To this end, we leveraged principles of peer production, a form of distributed knowledge production that shares characteristics with some types of citizen science, and that has been successfully applied in other areas of knowledge production to involve volunteers in tasks of varying complexity.

We developed a working model of peer production that encompasses hallmarks of knowledge production in peer production environments. Finally, we applied this model to three case studies to explore the usefulness of the model for a citizen science context and to provide an overview of how these citizen science projects implement participation.

This approach has two main limitations: First, our case studies focus on the main platforms of the citizen science projects at a static point in time. Further participation may occur outside of these platforms, for example, on coding platforms such as GitHub, in chat systems, or at online and offline community events. Additionally, looking at the status quo also risks ignoring differences in platform design over time, including design approaches that have been tried and discarded. Future research could complement our work by analyzing a longitudinal sample of platforms or by conducting interviews with long-time community members. Second, peer production literature is not specific to citizen science and the ecosystems for scientific knowledge production might come with their own particularities and requirements, potentially impacting the applicability of general peer production principles. Despite this, we note the similarities that some forms of citizen science projects share with peer production projects, as well as the potential of peer production approaches to engage volunteers in tasks of varying complexity²⁰. Importantly, peer production approaches have proven successful in coordinating the work of volunteers to create knowledge in science-adjacent fields such as the creation of encyclopedia articles or software code²⁸.

In line with this, the working model of peer production allowed us to identify and characterize participation architectures of three citizen science case studies along our seven key characteristics of peer production. In the following, we will make use of this structured approach to compare the case studies with regards to each characteristic. This will allow to highlight

similarities and differences, and reflect benefits, challenges, and applicability of various design decisions in different citizen science contexts. The case studies were selected to represent active online citizen science communities across a wide spectrum of participation approaches. These differences become already apparent in the self-stated goals of the platforms: Zooniverse's objective is to enable research projects at scale by giving research teams the opportunity to obtain crowdsourced data annotations. iNaturalist's goal is to "connect people to nature", and to promote conservation and biodiversity science. Public Lab aims to support communities in addressing environmental justice issues with scientific methods and open technology. While Zooniverse's goals are predominantly epistemic, iNaturalist, and notably Public Lab explicitly combine epistemic with societal benefit goals.

Different implications that these goals have on design decisions for the platforms can be found along all peer production characteristics we investigated, starting with the **(1) common research objects**: In Zooniverse, they constitute individual databases that research teams provide for their projects to be annotated by volunteers. Each research team has full decision-making power and control over the datasets and annotation process, including the exclusive right to consult the full dataset for the whole duration of the annotation project. In iNaturalist, the commonly created biodiversity database is public, and in turn can be consulted and reused by anyone. Public Lab focuses largely on mutual support with regards to research methods, with knowledge objects such as protocols for experiences, all shared under open licenses.

For the **(2) range of tasks** Zooniverse focuses mostly on a single type of tasks, i.e. annotating images or videos provided by researchers, in line with the contributory, or crowdsourcing nature. iNaturalist broadens this participation approach, as participants can annotate data points but also provide their own. Public Lab has a less structured approach, providing a more open-ended list of tasks related to experimental design. Those choices have downstream consequences on **(3) granularity and modularity**: In Zooniverse and iNaturalist tasks are of similar small granularity, as participants provide or annotate one data point at a time, which is automatically added to a database, leading to a low cost of integration. On Public Lab, where tasks are mostly text-based, there is more different granularity of tasks, and their aggregation into modules requires more effort due to the less-structured nature of participation: Data aggregation through tagging can require manual effort to connect individual contributions into a larger body of

knowledge to make them findable. To reduce the effort, Public Lab uses automatic assignment of tags, when users create artifacts from links.

Differences in the aforementioned characteristics also reflect on the **(4) equipotential self-selection** of tasks: With a focus on crowdsourcing⁷, Zooniverse automatically assigns data points to users, according to randomization rules determined during project creation, meaning no self-selection beyond picking a project the user wants to contribute to. This setting is helpful for newcomers needing expert guidance, which is provided by the controlled provision of data, tutorials and help features in Zooniverse projects²⁹. iNaturalist opens up the task selection by allowing users to choose which observations they would like to annotate, as well as to add new observations to the database. This freedom of choice helps users filter the database for observations of their interest and expertise, as a certain level of domain knowledge is necessary for species identification³⁰. A downside is that this can possibly lead to the introduction of sampling bias: Di Cecco et al.¹⁴ identify spatial, temporal, and taxonomic biases issuing from user observation patterns, e.g. because users tend to specialize on some species or groups, rarely observe a species twice, record observations in their free time, and close to where they live.

We also find different **(5) quality control** approaches, often in relation to data visibility, that are the result of differences in task design and main research objects. Zooniverse's focus on structured crowdsourcing favors an arithmetic quality control that is hidden from its users to avoid biasing future responses¹². Instead, users are given the chance to discuss their annotated data points post-classification. As iNaturalist focuses on creating a public database, all annotations are visible to everyone, with some allowances to obscure data to protect privacy or endangered species. While this communal validation approach could open the process to biases, research has found it leads to reliable classification quality³¹, potentially as species identification requires a certain level of expertise, and because the visibility of all annotations enables other participants to raise issues when noticing wrong annotations. Public Lab also uses communal validation. Given the platform's open-endedness, quality control is also more open-ended, often through qualitative feedback in written form on methods with the goal to help people collect "good enough data"³².

Different levels of visibility and transparency also affect **(6) learning trajectories** and the transition from novice to advanced tasks: In Zooniverse, project creators provide tutorials for

participants to learn the necessary skills. Given that other people's annotations remain hidden, learning from other participants is not enabled within the main annotation interface. Nevertheless, Mugar et al. found that meta-information left on the talk pages of data points serve as "practice proxies" that help inform users about norms of practice and can stimulate further engagement in higher-level analyses¹². While iNaturalist provides tutorials, users can also learn from the existing classifications that are publicly available, providing an additional pathway to onboarding and learning. Public Lab's mission is to empower communities to advocate for environmental justice, so it explicitly targets newcomers and encourages them to learn how to conduct their own studies. Endorsing trial-and-error as a key part of research¹⁷, the platform incorporates features that encourage asking questions, sketching problems and solution approaches, and reproducing each other's work.

Features for **(7) direct and indirect coordination** are used, for example, to signal open work to contributors, and are often a function, and part of other collaborative processes, like quality control. In Zooniverse open work is signaled by project leads. Users do not interact during the main annotation process, and annotations of other users are hidden, to avoid bias. In iNaturalist, users are guided to open work by flags that signal that a new observation needs to be identified, or by calls for participation to upload specific types of observations. In Public Lab, users can add specific artifacts that signal that they have a question or want feedback or guidance with regards to research protocols. All platforms have features that allow contributors to discuss topics further, either as comment functionality for specific instances of artifacts, forums, and chat systems. Naturally, there are more coordination features in projects where users need to find and allocate work to each other, or where users need to find work for their specific expertise. Discussion features attached to specific artifacts and projects are essential where unstructured feedback is needed, e.g. for entire research protocols. But also in crowdsourcing settings, their usefulness for deeper engagement and to leverage the creativity of citizen scientists for scientific discoveries has been praised^{12,13,33}, while also wishing for some more structured tools for coordination to scale up¹³.

Comparing the case studies, we find that the working model of peer production supports the identification and structured analysis of key differences among approaches and implementations of participation in citizen science projects. Shirk et al.'s⁶ framework of public participation in scientific research classifies our three case studies into different categories: Zooniverse most closely fits the definition of the "contributory" category, iNaturalist falls between "contributory"

and “collaborative”, and Public Lab is most closely “collegial”. Our model allows moving beyond such classifications by helping to understand the different design elements that help implement participation in citizen science: Many of these differences are a result of high-level design decisions regarding the goals and common research object of a given project, which in turn influence the types of participation that are enabled. For example, while the closed-ended goals of Zooniverse projects limit participation options, this approach in turn enables the automated large-scale integration of user-provided annotations while avoiding bias. Given iNaturalist’s goal of providing a public resource, some trade-offs were implemented regarding the potential for bias, but also enable new pathways to contribution while maintaining low integration-cost. Lastly, Public Lab’s open-ended nature provides a large set of potential contributions, but at the cost of a less-streamlined process for data integration.

In conclusion, the working model of peer production proved not only helpful for investigating and reflecting on individual projects, but also for comparing between, and understanding trade-offs and benefits of different designs. Our case studies are all examples of successful citizen science projects: Applying the model highlights how they use different engagement strategies to create value in different ways. While crowdsourcing and peer production are often contrasted in literature^{19,20}, we propose to conceptualize it rather as opposite ends of a spectrum. Our results suggest that already adding some peer production elements can make a big difference in terms of volunteer engagement. In peer production settings, individuals can choose to contribute to a range of tasks of different complexity, starting with micro-tasks as in crowdsourcing, or complex, creative tasks. The working model of peer production is thought to be applicable to a wide range of citizen science platforms, in order to situate the project based on peer production characteristics, to allow project leads to identify areas in which they might or might not integrate more collaborative approaches, depending on the characteristics and goals of their project.

Further research should apply the model to more citizen science projects, which is especially interesting given the wide range and loose definition of citizen science, to see how it applies to other contexts. It also needs to be explored to what extent the model is beneficial for the design of new or redesign existing platforms, and how it impacts the achievement of epistemic or societal goals. Potentially, the model could be extended by a set of guidelines, which could also integrate governance and intellectual property dimensions. Research on online infrastructures where data and knowledge is collaboratively created is underserved in citizen science literature, but especially important in this field where collaboration and user engagement are a core

element, and can ultimately influence how scientific knowledge is created. Extending on prior approaches to classifying citizen science projects, the present study goes beyond sorting projects into categories, allowing to understand the different design elements that help implement participation in citizen science across a wide range of approaches in online settings, from crowdsourcing to open-ended collegial projects. In doing so, the working model of peer production provides a systematic approach to help citizen science practitioners reflect on volunteer participation in data- and knowledge production with online infrastructures in use, as well as to inform the design of future projects.

Experimental procedures

Resource availability

Lead contact

Katharina Kloppenborg is the lead contact for this study and can be contacted at katharina.kloppenborg@cri-paris.org.

Data and code availability

This study did not generate any new datasets.

Materials availability

This study did not generate any materials.

Method details

Method 1: Working model

Peer production, while most commonly known from open source software or Wikipedia, is also applied in commercial tools such as TripAdvisor or Yelp. What unifies these is the underlying production process, i.e. decentralized task execution in an online environment, and the spanning of a wide range of possible tasks of varying complexity, from simple routine tasks to complex creative tasks, from which contributors are free to choose²⁰.

We derived the working model of peer production based on recent peer production literature, notably Kostakis and Bauwens²³, and Benkler²⁰, by identifying characteristics specifically related to knowledge production. The working model of peer production covers characteristics specific to peer production, as well as aspects of knowledge production that are not unique to peer production, but necessary to develop an understanding of how knowledge is created in an online environment. For more detailed information on the characteristics and the background of the model, see *Supplementary Material A*.

Method 2: Sampling and case studies

To identify citizen science platforms that might be relevant for analysis, we scanned project databases (eu-citizen.science, citzenscience.gov, scistarter.org, p2pvalue.eu), issues from 2016 to 2020 from the journal “Citizen Science: Theory and Practice”³⁴, and projects analyzed in a survey on the European citizen science landscape by Hecker et al. (2018). To select the cases, we used purposive sampling. The following conditions were considered: Each case had to be a citizen science platform, with an active, possibly large user base, and with a major part of knowledge creation happening online on the platform. Availability of secondary literature, especially if related to the keyword “peer production” was also considered. Beyond that, we tried to select platforms that cover a diversity of participation approaches. To get insights on participation approaches a priori, the “About” sections of the cases were consulted, as well as secondary literature. In the dataset used in Hecker et al. (2018), the projects were already annotated with Shirk et al. (2012)’s levels of participation.

To conduct our analysis, we applied the peer production working model to the three cases by inspecting the respective online platforms, complemented by supplementary literature if available. Links to features on the online platforms are available in *Supplementary Material B*.

We created a user account for each platform and walked through menu points and interaction possibilities to derive schematic knowledge creation processes. In the following, each peer production element from the model was applied to the case with respect to that knowledge production process. If secondary literature on specific design decisions was available, this helped to inform the platform comparison and discussion of benefits and challenges of different approaches.

Acknowledgements

We would like to thank Gary Wolf, Gwen Ottinger and Roland van Dierendonck for contributing to an early brainstorming on the role of peer production in citizen science. We would also like to thank Benjamin Mako Hill, Vineet Pandey, Molly Clare Wilson, Ariel Lindner and Enric Senabre for their discussions and feedback around this project. Furthermore, we thank Fotis Psomopoulos and the Open Life Science team for guidance and support during the realization of a peer production design practices blog based on the case studies. We also thank Susanne Hecker for kindly providing us with a list of the level of participation in European citizen science projects.

Thanks to the Bettencourt Schueller Foundation long term partnership, K.K. and B.G.T. were supported by a CRI Research Fellowship to Bastian Greshake Tzovaras. M.P.B was supported by funding from Open Humans Foundation.

Author contributions

Conceptualization, K.K., M.P.B., and B.G.T.; Methodology, K.K., M.B.P., B.G.T.; Investigation, K.K., M.B.P., and B.G.T., Writing - Original Draft, K.K.; Writing - Review & Editing, K.K., M.P.B., and B.G.T., Visualization, K.K.; Supervision, B.G.T.; Funding Acquisition, B.G.T., and M.P.B.

Declaration of interests

The authors declare no competing interests.

References

1. Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., and Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience* 59, 977–984. 10.1525/bio.2009.59.11.9.
2. Strasser, B.J., Baudry, J., Mahr, D., Sanchez, G., and Tancoigne, E. (2018). “Citizen Science”? Rethinking Science and Public Participation. *Sci. Technol. Stud.*, 52–76. 10.23987/sts.60425.
3. Christine, D.I., and Thinyane, M. (2021). Citizen science as a data-based practice: A consideration of data justice. *Patterns* 2, 100224. 10.1016/j.patter.2021.100224.
4. Frickel, S., Gibbon, S., Howard, J., Kempner, J., Ottinger, G., and Hess, D.J. (2010). Undone Science: Charting Social Movement and Civil Society Challenges to Research Agenda Setting. *Sci. Technol. Hum. Values* 35, 444–473. 10.1177/0162243909345836.

5. Haklay, M., Fraisl, D., Greshake Tzovaras, B., Hecker, S., Gold, M., Hager, G., Ceccaroni, L., Kieslinger, B., Wehn, U., Woods, S., et al. (2021). Contours of citizen science: a vignette study. *R. Soc. Open Sci.* 8. <https://doi.org/10.1098/rsos.202108>.
6. Shirk, J.L., Ballard, H.L., Wilderman, C.C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B.V., Krasny, M.E., et al. (2012). Public Participation in Scientific Research: a Framework for Deliberate Design. *Ecol. Soc.* 7, 29. <http://dx.doi.org/10.5751/ES-04705-170229>.
7. Howe, J. (2006). The Rise of Crowdsourcing. *Wired*.
8. Chiaravalloti, R.M., Skarlatidou, A., Hoyte, S., Badia, M.M., Haklay, M., and Lewis, J. (2022). Extreme citizen science: Lessons learned from initiatives around the globe. *Conserv. Sci. Pract.* 4. 10.1111/csp2.577.
9. Mah, A. (2017). Environmental justice in the age of big data: challenging toxic blind spots of voice, speed, and expertise. *Environ. Sociol.* 3, 122–133. 10.1080/23251042.2016.1220849.
10. Lewis, D.M. (2020). Do-It-Yourself Artificial Pancreas System and the OpenAPS Movement. *Endocrinol. Metab. Clin. North Am.* 49, 203–213. 10.1016/j.ecl.2019.10.005.
11. West, J., and O'mahony, S. (2008). The Role of Participation Architecture in Growing Sponsored Open Source Communities. *Ind. Innov.* 15, 145–168. 10.1080/13662710801970142.
12. Mugar, G., Østerlund, C., Hassman, K.D., Crowston, K., and Jackson, C.B. (2014). Planet hunters and seafloor explorers: legitimate peripheral participation through practice proxies in online citizen science. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (ACM)*, pp. 109–119. 10.1145/2531602.2531721.
13. Jackson, C., Crowston, K., Østerlund, C., and Harandi, M. (2018). Folksonomies to Support Coordination and Coordination of Folksonomies. *Comput. Support. Coop. Work CSCW* 27, 647–678. 10.1007/s10606-018-9327-z.
14. Di Cecco, G.J., Barve, V., Belitz, M.W., Stucky, B.J., Guralnick, R.P., and Hurlbert, A.H. (2021). Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. *BioScience* 71, 1179–1188. 10.1093/biosci/biab093.
15. Tinati, R., Van Kleek, M., Simperl, E., Luczak-Rösch, M., Simpson, R., and Shadbolt, N. (2015). Designing for Citizen Data Analysis: A Cross-Sectional Case Study of a Multi-Domain Citizen Science Platform. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (ACM)*, pp. 4069–4078. 10.1145/2702123.2702420.
16. Yadav, P., and Darlington, J. (2016). Design Guidelines for the User-Centred Collaborative Citizen Science Platforms. *Hum. Comput.* 3, 205–211. 10.15346/hc.v3i1.15.
17. Breen, J., Dosemagen, S., Blair, D., and Barry, L. (2019). Public laboratory: Play and civic engagement. In *The Playful Citizen: Civic Engagement in a Mediatized Culture*, pp. 162–174.
18. Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (Yale University Press).
19. Mahr, D., and Dickel, S. (2019). Citizen science beyond invited participation: nineteenth century amateur naturalists, epistemic autonomy, and big data approaches avant la lettre. *Hist. Philos. Life Sci.* 41, 1–19. 10.1007/s40656-019-0280-z.
20. Benkler, Y. (2016). Peer production and cooperation. In *Handbook on the Economics of the Internet*, J. Bauer and M. Latzer, eds. (Edward Elgar Publishing), pp. 91–119. 10.4337/9780857939852.
21. Fuster Morell, M., Senabre Hidalgo, E., and Cigarini, A. (2021). A framework for assessing the commons qualities of citizen science: comparative analysis of five digital platforms.
22. de Sherbinin, A., Bowser, A., Chuang, T.-R., Cooper, C., Danielsen, F., Edmunds, R., Elias, P., Faustman, E., Hultquist, C., Mondardini, R., et al. (2021). The Critical Importance of

- Citizen Science Data. *Front. Clim.* 3, 650760. 10.3389/fclim.2021.650760.
23. Kostakis, V., and Bauwens, M. (2021). Grammar of Peer Production. In *The Handbook of Peer Production*, M. O'Neil, C. Pentzold, and S. Toupin, eds. (Wiley), pp. 21–32.
 24. Zooniverse.org (n.d.). Zooniverse. <https://www.zooniverse.org/>.
 25. iNaturalist <https://www.inaturalist.org>.
 26. Public Lab <https://publiclab.org>.
 27. What is GBIF? (2022). GBIF Glob. Biodivers. Inf. Facil. <https://www.gbif.org/what-is-gbif>.
 28. Benkler, Y., Shaw, A., and Hill, B.M. (2015). Peer Production: A Form of Collective Intelligence. In *Handbook of Collective Intelligence* (MIT Press), p. 33.
 29. Østerlund, C., Mugar, G., Jackson, C., and Crowston, K. Typologies of Learning in Open Online Collaborative Communities. In *International Reports on Socio-Informatics (IRSI), Proceedings of the CSCW 2016 – Workshop: Toward a Typology of Participation in Crowdwork*, pp. 31–41.
 30. McMullin, R.T., and Allen, J.L. (2022). An assessment of data accuracy and best practice recommendations for observations of lichens and other taxonomically difficult taxa on iNaturalist. *Botany* 100, 491–497. 10.1139/cjb-2021-0160.
 31. Hochmair, H.H., Scheffrahn, R.H., Basille, M., and Boone, M. (2020). Evaluating the data quality of iNaturalist termite records. *PLOS ONE* 15, e0226534. 10.1371/journal.pone.0226534.
 32. Gabrys, J., Pritchard, H., and Barratt, B. (2016). Just good enough data: Figuring data citizenships through air pollution sensing and data stories. *Big Data Soc.* 3, 205395171667967. 10.1177/2053951716679677.
 33. Lintott, C.J., Schawinski, K., Keel, W., van Arkel, H., Bennert, N., Edmondson, E., Thomas, D., Smith, D.J.B., Herbert, P.D., Jarvis, M.J., et al. (2009). Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo? *Mon. Not. R. Astron. Soc.* 399, 129–140. 10.1111/j.1365-2966.2009.15299.x.
 34. Citizen Science: Theory and Practice <https://theoryandpractice.citizenscienceassociation.org/>.