Goals and Tasks: Two Typologies of Citizen Science Projects

Andrea Wiggins
School of Information Studies
Syracuse University
337 Hinds Hall, Syracuse, NY 13244 USA
Email: awiggins@syr.edu

Kevin Crowston
School of Information Studies
Syracuse University
346 Hinds Hall, Syracuse, NY 13244 USA
Email: crowston@syr.edu

***Abstract*—Citizen science is a form of research collaboration involving members of the public in scientific research projects to address real-world problems. Often organized as a virtual collaboration, these projects are a type of open movement, with collective goals addressed through open participation in research tasks. We conducted a survey of citizen science projects to elicit multiple aspects of project design and operation. We then clustered projects based on the tasks performed by participants and on the project's stated goals. The clustering results group projects that show similarities along other dimensions, suggesting useful divisions of the projects.**

I. Introduction

Citizen science is a form of research collaboration involving members of the public in scientific research projects to address real-world problems [1]. Active engagement in scientific work differentiates citizen scientists from those in less active roles, such as providing computing resources for projects like SETI@home or participating as a subject in a research study. Citizen science is related to long-standing programs employing volunteer monitoring for natural resource management [2], and is often employed as a form of informal science education or outreach to promote public understanding of science [3]. Experience with this model of science shows that with thoughtful study design and under the right circumstances, citizen science can work on a massive scale, generating high quality data that lead to reliable, valid scientific outcomes as well as unexpected insights and innovations [4], [1]. Such success motivates scientists to explore how members of the public might contribute in their projects.

However, there is an enormous variety of projects that fit under the rubric of citizen science. Despite initial efforts toward typology development, the diversity is not as yet well understood, which makes it difficult to determine how or if one project's experiences will be relevant to another. Further complicating this picture, the phenomenon of citizen science has evolved over time. For example, a number of the citizen science projects that emerged over the past two decades place more emphasis on scientifically sound practices and measurable goals for public education [5]. Virtual modes of contribution make it possible for a broader audience to engage in scientific work, as in other open content projects. An increasing number and variety of citizen science projects are taking advantage of the affordances of technology to advance scientific research [6].

Our goal in this paper is to develop an empirical typology that provides a more detailed understanding of the characteristics and need of citizen science projects. We build upon prior work with a larger sample of empirical data provided by the projects themselves. The resulting typology complements prior related work by providing a classification of projects based on a more detailed set of organizational, participatory, and technological characteristics than utilized in prior research.

II. Related Work

Citizen science project share characteristics with other kinds of open communities: there are similarities to peer production, open data is relatively common and open participation is nearly universal. Nevertheless, there are substantial differences that make existing typologies for these related phenomena unsatisfactory for describing citizen science. However, research on these related phenomena is useful as background to our conceptualizing.

Citizen science projects that are entirely mediated by information and communication technologies (ICTs) are often considered a form of crowdsourcing applied to science. Crowdsourcing is an ill-defined but common term referring to a set of distributed production models that make an open call for contributions from a large, undefined network of people [7], [8].

Virtual citizen science is clearly a different way of organizing online contribution than has been previously analyzed in the literature [9]. Unlike most online communities that have been studied, these projects are not self-organizing [10], [11], [12], [13]. Citizen science does not represent peer production in the same sense as seen in prior work because the power structure of these projects is nearly always hierarchical. Furthermore, citizen science is not necessarily "open science," a term that refers to open source-like practices in formal scientific research settings. Many citizen science projects share data, but many do not make the full research process publicly viewable for comment and discussion.

The structure of tasks is very similar to those of peer production, however, and existing literature is helpful for understanding key aspects of virtual citizen science [14], [15]. By

978-0-7695-4525-7/12 $26.00 © 2012 IEEE
DOI 10.1109/HICSS.2012.295

Stage of Inquiry	Cooper et al. [21]	Wilderman [22]	Bonney et al. [23]
Define question	✓	✓	✓
Gather information			✓
Develop hypotheses			✓
Design study	✓	✓	✓
Data collection	✓	✓	✓
Analyze samples		✓	✓
Analyze data	✓		✓
Interpret data	✓	✓	✓
Draw conclusions	✓		✓
Disseminate results			✓
Discuss results & ask new questions			✓

TABLE I
VOLUNTEER INVOLVEMENT IN ENVIRONMENTAL SCIENCE PROJECTS.
✓= TASK INCLUDED IN MODEL.

Stage of Inquiry	*Contributory*	*Collaborative*	*Co-created*
Define question			■
Gather information			■
Develop hypotheses			■
Design study		☒	■
Data collection	■	■	■
Analyze samples		■	■
Analyze data	☒	■	■
Interpret data		☒	■
Draw conclusions		☒	■
Disseminate results	☒	☒	■
Discuss results & ask new questions			■

TABLE II
THREE PARTICIPATORY SCIENCE MODELS (FROM [23]).
■ = PUBLIC INCLUDED; ☒ = PUBLIC SOMETIMES INCLUDED.

contrast, the hierarchical form of most virtual citizen science likely creates a different sense of community with respect to authority, leadership, decision-making and sustainability [16]. Finally, there are strong similarities with respect to issues of motivation and progressive engagement that bear a striking resemblance to virtual communities or networks of practice [17], [18], albeit with scientists as overseers of the community's practices. These prior models from studies of online communities of practice provide insight into the design of tasks and technologies to support virtual citizen science communities.

A. Tasks in Citizen Science Projects

Many attempts to describe citizen science projects have based their analysis on the tasks performed by the citizen participants. Table I lists the different steps in scientific inquiry that have been considered as part of citizen science in prior research. The forms of participation include contributing data according to an established protocol, or completing structured recognition, classification, or problem-solving tasks that depend on human competencies [19], [20]. Besides evaluating the stages of scientific inquiry in which the public is involved, Cooper et al. [21] included additional details of research, education, and management goals, which are contrasted in a framework for integrating individuals in monitoring and active conservation efforts in residential areas.

B. Typologies of Citizen Science Participation

Before presenting our own work, we review earlier typologies of citizen science projects. Typologies to date have focused primarily on the integration of public participation in different steps of scientific research, with little attention to sociotechnical and macrostructural factors influencing the design of the study or management of participation. These typologies examine the participation of the public by focusing on engagement in different steps of scientific research. The level of detail in these analyses differs, as do their final categorizations, yet they are largely in alignment.

Distinguishing *community science* from *citizen science* based on the community control of the inquiry, Wilderman [22] proposes an alternate typology that includes community consulting, community-defined research, community workers, and community-based participatory research. These categories are congruent with those presented elsewhere with some relatively small variations, and the author differentiates between two forms of community workers models based on whether or not analysis activities are exclusive to scientists, also adding a category in which the community is engaged in a consultative capacity, represented as "science for the people". Contrasting this practice against "science by the people" casts the typical scientist-initiated project model in a negative light; a more neutral perspective might suggest "science with the people" as another potential characterization.

A comprehensive, educationally-focused technical report that summarizes many of these views discusses *contributory*, *collaborative*, and *co-created* projects, shown in Table II, synthesizing many of the prior typologies [23]. The authors examined case study projects with a rubric-based evaluation to make a multi-faceted assessment of outcomes in several key focus areas. The final framework resembles a simpler variation on the other models, but includes more detail with respect to the steps of scientific inquiry in which volunteers may be included, moving the sophistication of the typology up a level despite its apparent simplicity.

These and other prior efforts targeted at understanding the defining features of citizen science projects have focused almost exclusively on the types of scientific tasks performed by volunteers, leading to a limited though functional view of participation. In contrast, [24] analyzed manually-collected data about citizen science projects to develop a typology with five types of projects—Action, Conservation, Investigation, Virtual, and Education—that differ in primary project goals and the importance of physical environment to participation. Action-oriented citizen science projects encourage participant intervention in local concerns, using scientific research as a tool to support civic agendas. Conservation projects support stewardship and natural resource management goals, primarily in the area of ecology; they engage citizens as a matter of practicality and outreach. Investigation projects are focused on scientific research goals requiring data collection from the physical environment; these projects best fit the definition of citizen science from [1]. In the science-oriented Virtual projects, all project activities are ICT-mediated with no physical elements whatsoever, differentiating them from the Investigation projects in which the physical places of volunteer participation was also important. Finally, the Education projects make education and outreach primary goals, all of which include relevant aspects of place. Building on this work, we examine a much broader range of characteristics to develop a typology focused instead on project goals and uses of technology to overcome the limitations of virtuality, which can help inform the development of cyberinfrastructure to support citizen science.

III. Methods

In this section, we describe the process we used to develop an empirically-grounded typology of citizen science projects. The study was designed as a survey of citizen science projects, eliciting project characteristics across eight categories of items. Our sample was drawn from email lists of existing citizen science projects and delivered as a two-part web-based questionnaire. The responses were analyzed with a hierarchical clustering algorithm, and we examined two different ways to cluster the projects, by participation tasks and by goals.

A. Instrument Design

A survey instrument was composed to directly elicit selected descriptive characteristics of projects, providing more specific and detailed data than was previously manually compiled from Internet sources and interviews in the prior study. It was presented as a two-part questionnaire: first, a brief project profile and second, a separate, lengthier survey.

The first portion of the questionnaire was a project profile, allowing projects to opt-in for listing on several cooperating websites that provide listings of citizen science projects, and update existing project profiles based on data provided with the sampling frame or create a new project profile. The project profile included 23 items for information that would be considered useful by potential participants, including project contact information and description, scientific domain, target audiences, geographic scope, participation locations, project duration, availability of learning materials, training requirements and required gear. Most items were multiple choice, with the options generated from the prior records about the existing projects, provided with the sampling frame. The remaining items, such as geographic scope and training requirements, were free response

The second portion of the questionnaire was the project survey, which asked for additional details in several categories: measures of project size (e.g., budget, personnel), project resources, features of participation, technology, data management, outcomes, goals, and evaluation. The survey included 57 items on 8 pages, including free-response spaces for each multiple choice or ranking matrix item to allow participants to respond more fully to the topics as desired. Each page represented a category of questions, with three or four multiple choice items and associated free response fields for most categories. There were no required fields, so each item had a variable response rate.

Under the category of project resources, we asked about project staffing levels, annual operating budget, and funding sources. To better understand participation details, we asked about types of participation activities, explicit rewards to contributors, and opportunities for social interaction. In the area of tools and technologies, we inquired about communication tools, technology development planned in the next two years, and areas of interest for future technologies. Data management items asked about data validation methods, data sharing practices, and data ownership policies. Under the category of project contributions, respondents defined the unit of contribution for their project, reported numbers of participant registrations, numbers of contributors, and number of contributions. Project goals and outcomes were addressed by ranking the importance of several project goal areas, and multiple choice items asking about intended and actual project outcomes, as well as evaluation practices. Finally, interest in participant outcome evaluation was elicited in the areas of science knowledge, science interest, science skills, attitudes, and behaviors.

B. Sample

The sampling frame was composed of projects listed on Cornell Lab of Ornithology's citizen science email list and in the now-defunct Canadian Citizen Science Network. These are the most comprehensive sources of contacts for the North American citizen science project leader community. Approximately 60 additional contacts were manually mined from the online community directory at http://www.scienceforcitizens.net to extend the disciplinary diversity of the sample.

These sources provided a combined set of approximately 840 contacts after removing duplicates and bad addresses. These contacts are individuals who had self-identified as responsible for or interested in the management of citizen science projects. Approximately 280 projects were identified in this process, and another 560 individuals who may be connected with additional projects were also invited to participate.

C. Survey Administration

We implemented our two-stage survey on a Drupal open source content management system using the Webforms module to provide basic survey functionality. The form permitted respondents to save drafts and return later to complete the survey. Each contact from our sample was used to create a site login. For each of the projects for which prior data were available from our collaborators (approximately 280), each project profile was seeded with the available data, e.g., contact information, organizational affiliations, topic keywords, and descriptive text. For contacts not previously associated with a project (approximately 560), only a login was created. These participants could create a project profile starting from scratch as well as filling out the survey.

Each contact received an advance message from the list manager at the Cornell Lab of Ornithology (a trusted authority), notifying recipients of the upcoming survey invitation and providing endorsement of the project. This was followed by an email invitation to respond in a message containing site login credentials. Two reminders were issued to contacts that had not yet updated or created a profile, at two weeks and five weeks after the initial contact.

D. Response Rate

In response to approximately 840 emailed requests for participation, 128 project profiles were created or updated. 73 surveys were initiated, of which 63 were fully completed. Three additional draft surveys provided enough responses to be considered usable (these respondents did not respond to the final two pages of the survey, which are not central to the analysis presented here), providing a total of 66 surveys for analysis. The surveys and profiles were combined for analysis of these projects.

The response rate is when compared to the number of contacts is about 8%, low, though not atypical for such a survey. However, it should be noted that contacts were asked to report on projects and the number of projects is smaller than the number of contacts. The Citizen Science Central site lists approximately 120 projects and Science For Citizens includes over 150 projects, with new projects surfacing regularly and we had initial profile data for approximately 280 projects. We estimate that this number is close to the size of the population (though it is growing quickly), meaning that our response rate provides acceptable coverage of projects.

The sample does have some limitations. Most of the known largest projects (in terms of funding, scale, and personnel) did not respond, and there were fewer responses from very small local projects than are suspected to occur in the larger population. Most of the responses came from small-to-medium sized projects, based in the United States, with several Canadian projects reporting and two from the UK. The sample therefore best represents North American citizen science projects. However, despite these limitations, we believe that the resulting sample is generally representative of the population of citizen science projects, as independent expert review of the response pool characteristics suggests that the responses provide a fairly representative sample of the larger community.

E. Missing Data

The survey included no required items and there are missing responses for most items. The per-item response rate varied; items presented later in the survey were answered less often, as were those eliciting specific quantities rather than Likert scales (e.g., annual budget, number of contributions, number of participants). Disregarding the free response items provided to allow further commentary, the lowest response rates were for the items inquiring into number of project registrations and contributions, both to-date and for 2010. Although 58 projects were able to define the unit of contribution for their project, the subsequent items received between 35 and 44 responses, with an additional 24 free responses elaborating on the challenges of quantifying participation. We expect that the low response on these items is because the data about participation were not readily available or accessible to respondents, as interviews have suggested is common for many mid-sized and smaller projects.

Outside of this category of questions, the lowest response for a single item was only 56 approximations of annual project budget. In addition to general sensitivity over financials, low response is most likely because the figures were not readily at hand for the responding individual, or because the project budget is not readily summarized due to integrating multiple (unquantified) resources from different organizations to enable project operations.

F. Data Analysis

The goal of our analysis was to develop an empirically-based typology of citizen science projects. We used statistical clustering methods to identify groups of similar projects, employing the R statistical software with the cluster and stats packages. Ward's minimum variance method, used with a euclidean distance measure, was chosen for clustering as it created fewer clusters with very small memberships. Based on our prior work and existing typologies of citizen science projects, we tested two alternate sets of variables for clustering. The first grouping was based on the scientific work tasks that comprise participation in the project. The second set of clusters was based on the relative emphasis on each of ten project goal areas. Because there was no natural number of clusters for the analyses (verified by comparing the sum of squares for different numbers of clusters), we chose to create five clusters in each case, as this resulted in the lowest incidence of clusters with only one or two members.

IV. Findings

In this section, we first provide descriptive statistics for the projects in our sample. We then examine the project clusters yielded by clustering on participation tasks, and alternately, project goals. At the same time, we examine the relationships of these clusters to other project characteristics included among the survey items.

A. Sample Description

The responding projects reported between zero and over 50 paid full-time equivalent employees (FTEs). Of 50 respondents for this item, the average number of staff was 2.4 but the median was one. Several noted that this allocation of staffing was spread across numerous individuals, each contributing only a small fraction of their time. Annual budgets ranged from $125 to $1,000,000 (USD or equivalent); 43 projects responded with estimated annual budgets, with an average of $105,000 but with a median of $35,000 and a mode of $20,000.

52 projects included the year founded in their responses. Responding projects were widely variable with respect to the age or duration of the project. A few projects were not yet operational, and one was 100 years old. The average age of currently operational projects is 13 years, while the median is 9 years and the mode is 2 years.

For 63 responding projects, funding came from four primary sources: federal grants (34, 54%), other grants (33, 52%), in-kind contributions (30, 48%, typically staff time) and private donations (20, 32%). While only a few projects collected revenues from membership fees (4, 6%), merchandise sales (4, 6%), sponsorships (3, 5%), or consulting and service fees (2, 3%), 11 (17%) collect fees from participants. Additional and more specific funding sources named included state appropriations, private foundations, government agencies. Projects employed up to five different funding sources to meet their expenses; however, several projects reported that they currently operate unfunded, with several comments suggesting that startup funding is easier to acquire than support for ongoing operations.

B. Participation Task Clusters

The majority of typologies of citizen science in the extant literature categorize the projects according to public involvement in different steps of the research process. Following this lead, we first clustered projects on this dimension. We focused on specific task types for which contributors might need training or prior experience. As a basis for developing the clusters, we used reports for the tasks identified in our prior work, which involved more detailed tasks than simply "collect data", as the different features of these tasks suggest a need for different skill or knowledge for effective participation. However, we chose not to include the full range of scientific tasks present in the typology from [23] because our prior work found a low incidence of public engagement in research tasks outside of data collection, management and analysis. The tasks and the associated clusters (labelled with a number) are shown in Table III below.

Additional activities that were reported as free text focus mainly on scientific tasks related to specific project requirements. These participant activities included:

Participation Tasks	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Observation	⊠	■	⊠	⊠	■
Species identification	⊡	⊠	⊠	⊠	■
Classification or tagging	⊡	□	⊡	⊡	□
Data entry	⊡	■	⊠	■	■
Finding entities	⊡	⊡	⊡	⊠	□
Measurement	⊡	⊡	⊠	⊠	□
Specimen/sample collection	⊡	□	⊡	⊠	□
Sample analysis	⊡	□	□	⊠	□
Site selection &/or description	⊡	□	⊡	⊠	⊡
Geolocation	□	□	⊠	⊠	□
Photography	□	□	⊠	⊠	⊡
Data analysis	□	■	⊡	⊠	□
Number of tasks	2–4	4–5	3–9	5–12	3–5
Projects in cluster	13	6	17	15	12

TABLE III
PARTICIPATION TASKS FOR EACH CLUSTER. ■=ALWAYS; ⊠=FREQUENTLY; ⊡=INFREQUENTLY; □=NEVER.

- Posing new questions, lit reviews, paper writing
- Videography
- Monitoring
- Insect rearing
- Organization and landowner coordination
- Identifying animal tracks
- Manual labor, habitat construction, shell recycling
- Creating maps
- Communication with other participants and scientists
- Sharing findings at meetings of related groups

The diversity of these additional participation activities clearly demonstrates that the nature of the contribution that participants make can varies substantially from one project to the next.

1) Description of Clusters: Each cluster shows a unique combination of task structures. For Cluster 1, most projects involve observation and identification tasks, but never require analysis. The members of this cluster represent "traditional" volunteer monitoring projects, focusing either on species observations or water quality, and involve the fewest participation tasks that were included on our questionnaire. In Cluster 2, all projects involve observation, data entry, and analysis, but include no locational tasks such as site selection or geolocation.

Cluster 3 projects engage participants in a variety of tasks; the only participation task not represented is sample analysis, indicating that there are no water quality monitoring projects in this cluster. Cluster 4 includes projects that engage the public in every participation task we considered, with fairly frequent incidence of tasks that suggest participants often select, describe, and document their own observation or collection sites. The projects in this cluster involve participants more intensively, permitting engagement in the largest number of tasks represented in this sample.

Finally, Cluster 5 projects all involve reporting, using observations, species identification, and data entry, but with very few additional participation opportunities. Clusters 2 and 5 appear very similar, with observation and data entry as common cornerstones. These projects never involve participants in classification tasks, geolocation tasks, or sample-related tasks, excluding the usual activities of sample collection and analysis in water quality monitoring projects. The primary difference between these two groups is the engagement of participants in data analysis: Cluster 2 projects all involve data analysis, while Cluster 5 projects do not.

Clusters 3 and 4 are also very similar, and the projects in Cluster 4 appear to be a more intensive variation on those in Cluster 3. Most of the tasks that occur infrequently among the Cluster 3 projects are very frequently found in Cluster 4 projects, and both share a strong incidence of observation, species identification, data entry, measurement, geolocation, and photography tasks. Cluster 4 projects involve participants more deeply; for example, sample or specimen collection is a very common task for Cluster 4 projects, but only occurs twice among the Cluster 3 projects. Cluster 1 projects represent a middle point between the observation and data entry model of Clusters 2 and 5, and the broad array of participation options in Clusters 3 and 4.

2) Relationships to Other Factors: We examined these clusters for relationships to other factors in the survey using Chi-square tests. Although there were a few weak relationships, we found only three strongly significant relationships to other factors. First, there was a relationship to the "no rewards" answer to the survey item inquiring about explicit rewards for participation (p = 0.003). Half the projects in Cluster 5 provide no rewards to participants, representing the majority of projects claiming to provide no incentives.

The second relationship was to the domain topic of technology (p = 0.009). Clusters 2 and 3 include all but one of the projects that identified technology as one of the topics of the project, while Clusters 1 and 5 have none at all. The third relationship we observed was with an ongoing, year-round duration of project activities (p = 0.009) as opposed to event-based, seasonal, or other durations. The majority of the projects in Clusters 1 and 5 have an ongoing year-round project duration, while Clusters 2, 3, and 4 include no more than two such projects.

C. Goal Clusters

As a second approach to developing a typology, we clustered projects according to their relative goal emphasis. On the survey, projects rated each of 10 goals (presented in random order) on a 7-point Likert-like scale, ranging from "not important" to "extremely important". Some projects marked every goal as somewhat or very important, while others were more discriminating. For analysis, we scaled each goal as a proportion of the whole, such that if the project marked every goal equally (e.g., all as "very important" or any other ranking), then each goal was weighted at 10%, while if only a single goal was ranked as "very important" and the others as "not important", then that goal would be weighted 100%.

Free responses listing additional goals included a variety of goals that were primarily aligned with outreach or education. These goals included (quotations indicate verbatim responses):

- Data appropriate for specific purposes (2)
- "Scientific literacy"
- "Get people outdoors and enjoying nature"
- "Get the public thinking about the process of biological evolution"
- "Spread ideas through a social network"
- "Positive youth development"
- "Enhance and support local volunteer programs"
- "Greater involvement in science, enhanced connection and feeling of ownership of environment"
- "Good science with no agendas to contaminate the findings"

1) Description of Clusters: The clustering algorithm produced five clusters, which we describe in terms of the relative emphasis the projects collectively placed on each goal, shown in Table IV (zeroes represent "not applicable"). Cluster A projects afforded nearly equal weights to each of the goal areas, and every project in this cluster rated every goal at the midpoint of the scale or higher. Cluster A projects also had the lowest overall nonzero ratings for science, while monitoring and stewardship were slightly more important on average. By contrast, Cluster B projects are most strongly focused on science, with the highest overall ratings for science among the projects, and very low ratings for management, action, and restoration goals. Cluster B projects also had the highest ratings for education and monitoring relative to the other clusters, as well as a strong emphasis on outreach.

In Cluster C, science is the most important goal, but education, monitoring, and discovery are only slightly less important on average. For Cluster D projects, science, conservation, monitoring and stewardship are most important, while discovery is less valued than in the preceding clusters. Cluster E contains only one project, the sole respondent giving a "not important" rating to science as a project goal, and rated an additional four areas as "not important." With every method for clustering on goals, this project was singled out as being significantly different from the others in the sample.

2) Relationships to Other Factors: Compared to the clusters generated by examining participation tasks, a larger number of relationships were identified with Chi-square tests, suggesting a more useful classification.

Although annual operating budget (p = 0.009) was a significant factor, several outliers make the patterns difficult to identify. If we exclude the two projects with million-dollar budgets, Clusters A and C have modest average budgets of $46,000 and $34,000, respectively. Likewise, the solo project in Cluster E has an annual operating budget of $35,000. Clusters B and D, however, have average budgets of $122,000, which demonstrates a substantial difference in available resources for these projects as compared to the other three clusters. Notably,

Goal	Cluster A	Cluster B	Cluster C	Cluster D	Cluster E
Science	0.10	0.17	0.13	0.16	0.00
Management	0.09	0.01	0.08	0.10	0.21
Action	0.10	0.01	0.09	0.10	0.00
Education	0.10	0.16	0.12	0.10	0.00
Conservation	0.10	0.09	0.10	0.12	0.00
Monitoring	0.11	0.16	0.12	0.14	0.21
Restoration	0.09	0.01	0.05	0.06	0.21
Outreach	0.10	0.14	0.11	0.08	0.17
Stewardship	0.11	0.11	0.09	0.12	0.21
Discovery	0.09	0.15	0.12	0.02	0.00
Projects in cluster	30	8	17	7	1

TABLE IV
AVERAGE WEIGHT GIVEN TO THE GOALS FOR EACH CLUSTER.

the two clusters with the highest emphasis on science as a project goal have the highest average budgets.

Another significant factor was training as an opportunity for socialization (p = 0.002). Positive responses to this option on the item asking projects to identify potential sources of social interaction suggest that collocated or at least synchronous training sessions are an option for some, if not all, participants in these projects. This was the most common case for projects in Clusters A, D, and E, but is infrequent for Cluster C projects and not available for Cluster B. The primary difference between the science-focused projects in Cluster B and Cluster D in this respect is that the projects in Cluster D tend to be more geographically constrained, making in-person trainings more practical.

Interestingly, another related factor, the availability of online training materials (p = 0.002), further reinforces this split. While we might expect that the projects without face-to-face training opportunities would compensate with a greater likelihood of online training materials, this was not the case. Clusters A, D, and E, which noted the opportunity for socializing through training sessions, also reported providing online training materials with greatest frequency, at over three quarters of projects in each cluster. While projects in Clusters B and C were not particularly remiss in this respect, there was a slightly higher incidence of projects without training materials, and there was no direct substitution effect of online training materials in place of face-to-face training. This suggests that these projects are designed to require minimal training or may rely on existing skills.

A third potentially related factor, the use of websites as communication tools (p = 0.006), similarly distinguishes the clusters along these lines. While only five projects out of 63 did not report using a website, none of them fell into Clusters B and C. This indicates that projects with face-to-face interaction opportunities are slightly less likely to rely on web-based materials for communication with participants.

A similar pattern emerges for projects reporting that data ownership policies were currently under development (p = 0.002); only 4 project indicated that this was the case, and none of those projects are members of Clusters B and D. Although this might seem to suggest that the two clusters with the strongest emphasis on scientific goals have established data ownership policies, this was not the case, as several of the projects in these clusters (as in the others) have no such policy.

Projects focusing on adults as target audience (p = 0.008) were well-represented; an answer of "all of the above" on the item asking projects to identify target audiences was interpreted as a positive answer for all of the other items. Again, the projects that did not indicate that adults are a target audience were in the minority, with only five such projects in the sample. Notably, however, some of the audiences are overlapping, e.g., engaging community groups or landowners most likely targets primarily adult participants. If we take this perspective, there are only two projects, both in Cluster A, that do not focus on adults as a primary audience. Both of these projects marked schools as primary audiences, which we interpret to mean that schoolchildren are the primary participants.

Examining action as an expected project outcome (p = 0.005) provided slightly less distinctive patterns. Clusters A, C, and D had approximately 50% of the projects responding that they intended to promote action by contributors. The project in Cluster E did not, and only one of the eight projects in Cluster B included action as an expected outcome. This clearly suggests that the projects most strongly focused on scientific goals had the least interest in promoting action outcomes from participants, which is in keeping with the low rating of action as a goal for these projects.

V. DISCUSSION

In this section, we discuss the conceptual implications of the analysis and future work.

A. *Comparing Clustering*

To compare the results of our prior work on typology development based on projects goals to the existing literature that develops typologies according to participation tasks, we performed algorithmic clustering according to both of these sets of characteristics. Clustering on participation, while intuitive and straightforward, does not highlight conceptually interesting relationships. What it does show is that some projects use more technologies to support participation tasks, that some participation structures naturally support ongoing, year-round activities, and that some projects choose to rely on altruism as a participant motivation. These are fairly obvious relationships with straightforward interpretations, providing relatively little room for conceptual development.

By contrast, when we clustered projects by their goals, more interesting patterns emerged, providing a richer conceptual basis for understanding citizen science projects. There are clear similarities between the science-oriented Clusters B and D, but also differences, as these two Clusters focus differentially on the education and outreach versus conservation and action goals, while also differing in geographic scale.

Cluster D's focus on conservation and action is linked to more regional and local projects, with more face-to-face interaction opportunities, particularly through in-person training events. At the same time that Cluster B's larger geographic scope (national to global) makes in-person training impossible, there is also a lower frequency of providing online training materials, suggesting that the participation tasks for this cluster have been designed for simplicity. These links between geographic scale and structure of participation are intuitive, given the coordination requirements for each type of project.

Besides training materials, however, we were surprised to see no relationship emerge between education goals and any other types of education materials. This suggests the primary educational benefit to participants is gained through participation. A lack of alignment between goals and participation tasks suggests that there is no set of tasks that indicates a project's goal orientation. Further, there was no relationship between participation clusters and much of anything else. This lack of relationship is reflective of the utility of grouping projects by the allocation of participation tasks between experts and non-experts, as opposed to creating a typology that takes on goals, which we find are aligned with other organizational characteristics.

B. *Implications*

The relationship of goals with geographic scale and in-person versus independent social experiences suggests that the larger scale projects are indeed more virtual. These project have a different set of requirements for technology and social support than projects that occur at more localized geographic scales. The more virtual projects may need more explicit social technologies, such as forums, blogs, and social media, to generate the social benefits to the project participants that were previously provided by in-person training events. The participation tasks are also more likely to be designed for minimal training when the opportunity to provide in-person support is removed by distance. This suggests that data verification will be more of a concern as well, since there is less training and no opportunity to observe the participant's performance in a data collection situation. On the other hand, for projects reliant on participants' existing skills, the investment may be in volunteer management rather than training materials, as recruitment of participants with the right experience is often a more intensive effort.

We also saw that projects with a stronger emphasis on science had larger budgets. Many of these project had access to grant funding, whether at the federal or state level, which could support an increased project capacity in terms of staffing and data management. This implies that small shoestring-budget projects may suffer in scientific quality due to the lack of resources to help coordinate the additional efforts that support data quality. However, as the difference in operating budgets between our clusters was relatively small (after the most extreme outliers were removed) this suggests that with relatively little additional investment, such as resources to support an additional staff person, a substantial improvement in project outcomes could be achieved.

The lack of data ownership policies among some of the more strongly science-oriented projects was unexpected. We noted that the largest projects (by any measure) all had data ownership policies, and it's possible that smaller projects may not be fully aware of data management, ownership, and intellectual property concerns. These issues may also be more carefully monitored in projects run by government agencies, as related policies such as the US Paperwork Reduction Act create a different set of constraints for enlisting public participation.

Scientific collaborations with thousands of contributors have previously occurred most often in scientific domains that are heavily reliant on major infrastructure (e.g., the Large Hadron Collider), conventions have been established because all parties have a vested interest in receiving credit for their work. In citizen science projects, by comparison, project organizers are more likely to come from domains where small team collaboration is most common and many contributors may be largely disinterested in this aspect of research. This does not eliminate the potential legal and ethical considerations of scientific collaboration, but it is not clear what consequences may arise for projects that do not have established policies regarding data ownership.

C. *Future Work*

Extending this work in the future with a more comprehensive sample would provide a clearer view of the full range of citizen science project diversity. In particular, there is an opportunity to collect data from projects outside North America. While citizen science project organizers are relatively well networked in North America, there is little communication across continents. Although there are known active citizen science communities in Europe, there are as yet fewer project directories that can be leveraged to identify these projects for sampling. This challenge is compounded by the issues of multilingual searching, and the current lack of established terminology to refer to the phenomenon of citizen science in non-English languages (although the issue of nomenclature certainly arises in English as well). As the global community of practitioners organizing citizen science projects continues to develop, future work can expand on these findings by further research employing more internationally representative samples.

Additional data points that could shed light on some of the differences between projects in the goal-based clusters would include the number and type of institutional links. Whether the project is run by a single academic PI or by a collaborative arrangement between multiple conservation organizations can be expected to have a significant influence on its resources, impacting the geographic range and types of goals the project undertakes.

In addition, further variations of project features could be examined for usefulness in clustering project types for a descriptive typology that provides insight into project characteristics and needs. Geographic range and other aspects of the physical qualities of participation also seem to have potential to prove as enlightening as project goals with respect to the relationship of place to other project characteristics.

VI. Conclusion

Virtual citizen science represents a new type of open movement, welcoming contributions to scientific research from a diverse population of volunteers. This domain of practice is rapidly expanding with the availability of enabling technologies and mounting evidence in favor of the efficacy of this research strategy. Prior citizen science typologies have focused primarily on the integration of public participation in different steps of scientific research, and this paper complements the prior work with a goals-oriented clustering of project that suggests that emphasis on science, along with the level of emphasis on education versus conservation, provides a useful way of distinguishing between projects. Our goal in developing this typology was to advance our understanding of the characteristics and needs of citizen science projects. Our findings indicate a relationship between resources, geographic scale, and the relative emphasis on different combinations of goals in citizen science projects. These differentiations also suggest that technologies to support projects with different goals may need to provide social support as well as task support as projects grow in scale.

Acknowledgments

The authors thank Jennifer Shirk, Tina Phillips, and Rick Bonney of the Cornell Lab of Ornithology, and Marlene Doyle of Environment Canada, for their assistance with project sampling. This research has been funded by US NSF OCI Grant 09-43049.

References

[1] J. P. Cohn, "Citizen Science: Can Volunteers Do Real Research?" *BioScience*, vol. 58, no. 3, pp. 192–107, March 2008 2008.

[2] K. Firehock and J. West, "A brief history of volunteer biological water monitoring using macroinvertebrates," *Journal of the North American Benthological Society*, vol. 14, no. 1, pp. 197–202, 1995.

[3] D. Brossard, B. Lewenstein, and R. Bonney, "Scientific knowledge and attitude change: The impact of a citizen science project," *International Journal of Science Education*, vol. 27, no. 9, pp. 1099–1121, 2005.

[4] D. Trumbull, R. Bonney, D. Bascom, and A. Cabral, "Thinking scientifically during participation in a citizen-science project," *Science Education*, vol. 84, no. 2, pp. 265–275, 2000.

[5] R. Bonney, C. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. Rosenberg, and J. Shirk, "Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy," *BioScience*, vol. 59, no. 11, pp. 977–984, 2009.

[6] J. Silvertown, "A new dawn for citizen science," *Trends in Ecology & Evolution*, vol. 24, pp. 467–471, 2009.

[7] J. Howe, "The rise of crowdsourcing," *Wired Magazine*, vol. 14, no. 6, pp. 1–4, 2006.

[8] D. Brabham, "Crowdsourcing as a model for problem solving: An introduction and cases," *Convergence*, vol. 14, no. 1, p. 75, 2008.

[9] D. Ellis, R. Oldridge, and A. Vasconcelos, "Community and virtual community," *Annual review of information science and technology*, vol. 38, pp. 145–188, 2004.

[10] M. Markus, B. Manville, and C. Agres, "What makes a virtual organization work?" *Sloan Management Review*, vol. 42, no. 1, pp. 13–26, 2000.

[11] M. Wasko, S. Faraj, and R. Teigland, "Collective action and knowledge contribution in electronic networks of practice," *Journal of the Association for Information Systems*, vol. 5, no. 11-12, pp. 493–513, 2004.

[12] S. Bryant, A. Forte, and A. Bruckman, "Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia," in *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work*. ACM, 2005, p. 10.

[13] K. Crowston, Q. Li, K. Wei, U. Eseryel, and J. Howison, "Self-organization of teams for free/libre open source software development," *Information and software technology*, vol. 49, no. 6, pp. 564–575, 2007.

[14] Y. Benkler, "Coase's Penguin, or, Linux and the Nature of the Firm." *Yale Law Journal*, vol. 112, no. 3, pp. 367–445, 2002.

[15] C. Haythornthwaite, "Crowds and Communities: Light and Heavyweight Models of Peer Production," in *Proceedings of the Hawai'i International conference on Systems Sciences, 2009*, 2009.

[16] B. Butler, "Membership size, communication activity, and sustainability: A resource-based model of online social structures," *Information Systems Research*, vol. 12, no. 4, pp. 346–362, 2001.

[17] J. Preece and B. Shneiderman, "The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation," *AIS Trans. on Hum.-Comp. Interact.*, vol. 1, no. 1, pp. 13–32, 2009.

[18] G. Fischer, "Beyond 'couch potatoes': From consumers to designers and active contributors," *First Monday*, vol. 7, no. 12-2, 2002.

[19] B. Sullivan, C. Wood, M. Iliff, R. Bonney, D. Fink, and S. Kelling, "eBird: A citizen-based bird observation network in the biological sciences," *Biological Conservation*, vol. 142, no. 10, pp. 2282–2292, 2009.

[20] A. Cho and D. Clery, "Astronomy Hits the Big Time," *Science*, vol. 323, no. 5912, p. 332, 2009.

[21] C. B. Cooper, J. Dickinson, T. Phillips, and R. Bonney, "Citizen Science as a Tool for Conservation in Residential Ecosystems," *Ecology and Society*, vol. 12, no. 2, 2007.

[22] C. C. Wilderman, "Models of community science: design lessons from the field," in *Citizen Science Toolkit Conference*, C. McEver, R. Bonney, J. Dickinson, S. Kelling, K. Rosenberg, and J. L. Shirk, Eds., Cornell Laboratory of Ornithology, Ithaca, NY, 2007.

[23] R. Bonney, H. Ballard, R. Jordan, E. McCallie, T. Phillips, J. Shirk, and

C. Wilderman, “Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report,” Center for Advancement of Informal Science Education (CAISE), Washington, DC, Tech. Rep., 2009.

[24] A. Wiggins and K. Crowston, “From conservation to crowdsourcing: A typology of citizen science,” in *Proceedings of the Forty-fourth Hawai'i International Conference on System Science (HICSS-44)*, Koloa, HI, 1/2011 2011.