Coordinating Advanced Crowd Work: Extending Citizen Science

Kevin Crowston  
Syracuse University  
crowston@syr.edu

Erica Michelle Mitchell  
Syracuse University  
emmitc01@syr.edu

Carsten Østerlund  
Syracuse University  
costerlu@syr.edu

Abstract

Crowdsourcing work with high levels of coupling between tasks poses challenges for coordination. This paper presents a study of two online citizen science projects that involved volunteers in such tasks: not just analyzing bulk data but also interpreting data and writing a paper for publication in one project and identifying new classes of data in the other. However, extending the reach of citizen science adds tasks with more dependencies, which calls for more elaborate coordination mechanisms but the relationship between the project and volunteers limits how work can be coordinated. Contrariwise, a mismatch between dependencies and available coordination mechanisms can be expected to lead to performance problems. The results of the study offer recommendations for design of citizen science projects for advanced tasks.

1 Introduction

The past decade has seen a rapid growth in the number of online citizen science projects. However, most of these projects involve the crowd in rather simple tasks (i.e., microtasking). Studies of citizen science volunteers suggest that many are motivated by the opportunity to contribute to real science (Raddick et al., 2010, Rotman et al., 2014) and by recognition for such contributions (Rotman et al., 2012). Accordingly, some sponsors of citizen science projects seek to involve
volunteers more deeply in the science of the project: not just collecting or processing data, but also taking part in further data analysis and even paper writing (Newman et al., 2012).

Efforts to further involve volunteers in more advanced tasks are also viewed as important in part to demonstrate that citizen science is not just crowdsourcing without pay, an exploitation of the citizen scientist volunteers by project scientists. To be fair to the volunteers, project scientists need to give back (Riesch and Potter, 2014, Owens, 2016), and expanding access to science is one way to do so. Allowing participants to see and talk about the data is only the first step in expanding access (Woodcock et al., 2017).

However, more advanced science tasks often comprise many interdependent pieces. To successfully include volunteers more deeply in scientific research requires careful consideration of the kind of project management needed, i.e., how to coordinate their contributions to complex tasks. This paper addresses the following research question: What coordination challenges do citizen science volunteers face when undertaking work with a high level of coupling?

To answer this question, this paper presents a case study of work in two projects: Galaxy Zoo Quench and in Gravity Spy. Quench was a project sponsored by the Zooniverse in which volunteers were invited to write an academic paper in collaboration with the project scientists. The Galaxy Zoo project had already had great success involving volunteers to work on classification of galaxies. The capability of the volunteers to do original work had seemingly been proven by discoveries such as Hanny’s Voorwerp, a novel astronomical object identified by a citizen scientist (Lintott et al., 2009). Furthermore, citizen science volunteers had been observed to engage in their own analyses of project data, posting questions and results to the discussion boards (Tinati et al., 2015, Bonney et al., 2009) and some had been involved individually in further re-
search. The next logical step appeared to be involving volunteers in scientific collaboration through the entire process of scientific research, from data analysis to publication.

Gravity Spy (Zevin et al., 2017) is another project hosted on the Zooniverse (Simpson et al., 2014) platform. The goal of Gravity Spy (www.gravityspy.org) is improving the instruments used to search for gravitational waves in the Laser Interferometer Gravitational-Wave Observatory (LIGO) scientific collaboration. The high sensitivity of the detectors needed to detect evidence of gravitational waves means that they are also extremely sensitive to noise, both external (e.g., earthquakes or vehicular traffic) and internal (e.g., parts of the instrument interacting in unexpected ways). When this noise is recorded by the interferometers (called a “glitch”), it potentially blocks detection of gravitational wave signals, so understanding glitches and removing their sources is a key activity to improve the sensitivity of the detector. Having collections of glitches of the same class is useful to the LIGO engineers as they seek to locate and remove the cause of the glitches. To that end, volunteers perform two tasks, classifying glitches into existing glitch classes and identifying possible new glitch classes (“advanced work”). We focus on the second task in this paper.

Theoretically, we draw on coordination theory to explore the challenges associated engaging members of a crowd in advanced science tasks. The two cases are rich and can be viewed from numerous perspectives, but we chose coordination theory for our analysis because it seemed to provide insight into the challenges faced by a distributed group trying to work together. To support this analysis, we first analyze the work of citizen science projects and the process of writing an article, to explore the nature of dependencies and coordination that would apply in the individual phases of the Quench project and to Gravity Spy advanced work.
2 Theory: Coordination theory

We first introduce the topic of coordination and present the fundamentals of coordination theory, the theoretical foundation for this paper. Coordination, defined as “managing dependencies between activities” (Malone and Crowston, 1994), is a central feature of collective action. This definition of coordination is consistent with the large body of literature developed in the field of organization theory (e.g., Galbraith, 1973, Thompson, 1967) that emphasizes the importance of interdependence in group work.

Coordination theory (Malone and Crowston, 1994) synthesizes contributions from different disciplines to develop a systematic approach to the study of coordination. Malone and Crowston (1994) analyzed group action in terms of actors performing interdependent tasks to achieve some goal; i.e., in an organizational process (Crowston, 1997, Crowston and Osborn, 2003). These tasks might require or create various resources. For example, in the case of writing a scientific paper, actors include the authors and various members of the research team. Tasks include collecting data, performing analyses and writing a revising a manuscript. Resources include data, analysis reports and the analysts’ and authors’ time and effort. In this view, actors in collective action face coordination problems arising from dependencies that constrain how tasks can be performed. Studying coordination thus means analyzing the dependencies that emerge among the tasks in a system and identifying how those dependencies are managed.

In contrast to other theories that consider dependencies among actors, coordination theory classifies dependencies as occurring between a task and a resource, among multiple tasks and a resource, and among a task and multiple resources. The dependencies between a task and a resource are shown in Figure 1. Dependencies between a task and a resource arise because a task uses or creates a resource. For example, a data analysis task uses data that has been collected and
preprocessed and creates analysis reports that might be used to write a paper. Resources may also be directly interdependent due to physical connections (the right side of Figure 1), e.g., a section of a paper that refers to results established in a prior section or data sets that need to be analyzed as an ensemble.

Shared use of resources can in turn lead to dependencies between the tasks that use or create the resource. These dependencies come in three kinds, as shown in Figure 2. First, producer-consumer or flow dependencies match Thompson’s sequential dependency (Thompson, 1967): one task creates a resource that a second uses. For example, in a data analysis pipeline, the flow of data from one analysis to another creates a dependency between those tasks. Flow dependencies further imply the need to manage the usability of the resource and the timing and location of its availability (that is, a flow dependency has three aspects), e.g., data from one stage of an analysis pipeline must be suitable for the next stage and made available on time.

Second, a shared-output or fit dependence occurs when two activities collaborate in the creation of an output (in the case where the output is identical, there is potential synergy, since the duplicate work can be avoided). For example, data analyses to support a paper need to be tai-
lored to work together.

Finally, a shared-input dependency emerges among activities that use of a common resource (like Thompson’s pooled dependency). For example, data collection might require a specific scientific instrument, constraining how data collection tasks are done, e.g., a schedule of observation times. Note that information as a resource is shareable, which can ease management of shared input dependencies, but simultaneously creates a different dependency of ensuring that different tasks are working with the same version of the data.

The key point in coordination theory is that dependencies create problems (or possible synergies) that may require additional work to manage. Malone and Crowston (1994) called this additional work coordination mechanisms. For example, if expertise is necessary to perform a given task (i.e., there is a task-actor dependency), then an actor with that expertise must be identified and the task assigned to him or her. The work of identifying an expert and maintaining a task assignment system constitutes the coordination mechanism.

There are often several mechanisms that can be used to manage a given dependency. For example, to manage a usability dependency (part of a flow dependency), the resource created might be tailored to the needs of the consumer (meaning that consumers must provide information about their needs to the producer) or a producer might follow a standard so the consumer knows what to expect. Usability dependencies are particularly salient in scientific research. Data that are collected must be appropriate for the research question and be credible according to the standards of the field. Analysis reports must meet the expectations of the field and provide answers to questions of interest. Papers must be written in the genre of a scientific paper, with the details of the genre differing from field to field. An important part of the training of a scientist is to learn the specific expectations for data, analysis reports and papers in the scientist’s research
field. That is, the expertise needed to do a task includes knowing how to do it in the way expected by users of the output.

It should be noted that in developing the coordination theory framework, Malone and Crowston (1994) describe coordination mechanisms as relying on other necessary group functions, such as decision making, communications and development of shared understandings and collective sense making (Crowston and Kammerer, 1998). To develop a complete model of a process would involve modeling all these aspects: coordination, decision making, communication and sense-making. In this paper though, we will focus on the coordination aspects, mostly bracketing the other phenomenon.

In summary, coordination theory provides a lens with which to analyze group processes in terms of tasks, resources, resulting dependencies and selected coordination mechanisms. Furthermore, the fit or lack of fit between the dependencies and available coordination mechanisms may explain problems faced by the group in achieving its goals.

2.1 Coordination in citizen-science projects and in paper writing

In this section, we present a theoretical analysis of citizen science projects from a coordination-theory perspective as a basis for analyzing the work of Galaxy Zoo Quench and Gravity Spy. We start by presenting an analysis of the work of Galaxy Zoo, which is a prototypical online citizen science project as well as the basis for the Quench project. This analysis is based on our own experience with the site and published studies of these citizen science projects (e.g., Prestopnik and Crowston, 2012, Wiggins and Crowston, 2015, Tinati et al., 2015, Simpson et al., 2014). The quality of the data created by the citizen scientists for scientific research emerges as a key issue from these analyses (Riesch and Potter, 2014, Wiggins et al., 2011) and provide a comparison point for understanding the more ambitious work of Galaxy Zoo Quench and the ad-
vanced work in Gravity Spy. We then develop an analysis of the coordination needed for the task of writing a paper, as writing a paper was the goal of the Quench project.

2.1.1 Galaxy Zoo. Galaxy Zoo (http://galaxyzoo.org/) is a citizen science project that has volunteers support scientific inquiry by online analysis of the millions of astronomical photographs collected by the Hubble Space Telescope, the Sloan Digital Sky Survey, and others. Specifically, the Galaxy Zoo system asks individuals to answer a series of questions about the shape of a galaxy captured in an image (e.g., the number of spiral arms or how round or elliptical they are). The resulting data supports astronomical research on galaxy morphology. The workflow for the data analysis task in the project, from galaxy classification to astronomical research, is shown in Figure 3. While we focus on describing Galaxy Zoo, this workflow describes many online data-analysis citizen science projects, including the main work of Gravity Spy.

Figure 3 shows the flow of data from occasional serendipitous discoveries. Every image is inspected by human analysts who may identify oddities in the images, such as the Voorwerp. As the figure shows, such discoveries are handled outside the regular flow in the project and support research other than the planned project research (Tinati et al., 2015).

One coordination problem in the Galaxy Zoo project is task assignment, matching an image to be classified to a volunteer. In the Galaxy Zoo project (and other Zooniverse projects), this dependency is handled by the system simply giving the next image to be classified to the next available volunteer who has not already seen it (Reeves et al., 2017). This approach has the advantage of being simple and requiring no information about the image or volunteer. A second problem is ensuring data quality, that is, the usability of the data classifications for the research project. In Galaxy Zoo (and similar projects), this usability dependency is handled by having multiple volunteers repeat the classification and taking the consensus.
In summary, the tasks of Galaxy Zoo have minimal dependencies that can easily be handled by the system. As a result, the level of coordination needed in the Galaxy Zoo and similar projects is minimal.

2.1.2 Paper writing. In contrast to citizen-science classification, the dependencies in writing a scientific paper are more complicated. Figure 4 shows the structure of dependencies involved, based on published work on coordination in writing (Erkens et al., 2005), Wikipedia in particular (e.g., Kittur et al., 2009) and a detailed coordination-theory analysis of a comparable process, writing software (Crowston and Scozzi, 2002, Crowston and Scozzi, 2008).

A first difference between Figures 3 and 4 is the presence of dependencies between the parts of the paper, the outputs of the paper writing tasks. Only a few tasks in writing, such as proofreading, are like galaxy classification in that they can be done without affecting other tasks (Kittur et al., 2009), i.e., by crowdsourcing (Bernstein et al., 2015). For the most part, different parts of a paper cannot be written independently. For example, the research problem presented in the introduction to a paper must be supported in the literature review, answered in the data analysis, and so on (Wichmann and Rummel, 2013). Furthermore, the voice and writing style of the different sections needs to match. These dependencies among parts of a paper impose constraints on how the paper parts are written (Kittur et al., 2009). To manage these dependencies requires

![Figure 3. Flow of data in the Galaxy Zoo project.](image1)

![Figure 4. Expected structure of dependencies in writing a paper.](image2)
additional work as authors must either plan the writing process in advance (Viégas et al., 2007, Erkens et al., 2005), e.g., by developing a shared vision for the paper (Wichmann and Rummel, 2013) (collectively or led by one person, Kim et al., 2014), or writing and revising their parts to fit with other parts. Teevan et al. (2016) report on a system to create microtasks to support paper writing, but despite the design intent, observed “considerable interaction among group members” using the system.

A second dependency is a shared-output dependency, created when two authors work on tasks that have the same output, i.e., two authors working on writing the same part of the paper. Galaxy Zoo also has multiple volunteers work on the same galaxy image, but because there are a small number of possible results, a simple consensus rule is usually sufficient to merge the classifications. However, many more differences can arise in writing a paper. At a basic level, problems of simultaneous changes to text can be managed by a shared document editor such as Google Docs (Lowry and Nunamaker, 2003). However, there can be problems at a conceptual level that are more difficult to identify and resolve (Erkens et al., 2005). To manage this dependency requires some technique to mitigate these possible conflicts in output, e.g., picking one version and rejecting the others or manually merging the changes.

A third dependency is the task-actor dependency. Unlike the system assignment in Galaxy Zoo, volunteers working on a paper will likely chose for themselves which tasks to work on, as in Wikipedia. Reliance on self-assignment of tasks fits the voluntary nature of the project but raises two potential problems.

First, people choosing to work on some part of the paper might not be good at it, i.e., their contributions might not be usable. In a conventional team, members would be assigned to tasks based on skills, but in a voluntary setting, skills are not guaranteed. A paper writing process
will have to include mechanisms to assess if a writing contribution is acceptable (Kittur et al., 2007). For example, in Wikipedia, editors police edits and modify or revert problematic ones. Conversely, efforts could be made to provide the volunteer with the necessary skills, e.g., by providing training.

Second, a volunteer might not be reliable, meaning that a promised contribution might not appear (Riesch and Potter, 2014). The writing process will thus also need mechanisms to handle missing contributions. This problem interacts with the second dependency, shared output, as one way to minimize problems from the former issue is to have only one person at a time work on a task (i.e., assign authors for each document section), but such a process is problematic if there is a chance that the task (i.e., the document section) will not be completed.

A final dependency is between the creation of the paper and the use of the paper by its intended audience. In the basic work of citizen science projects, the usability of the resulting data set is managed by having the science teams design the process of creating the data, with carefully imposed quality checks (Simpson et al., 2014). For scientific writing, this dependency is handled in part by processes such as peer review that check for article quality. However, much of the process is handled by the authors themselves acting as proxies for the readers. Knowing the scientific literature, scientific authors pick topics and write in ways that they know will be useful for that community (e.g., in the genre of a scientific article). A volunteer-driven writing process will need ways to provide information about the needs and desires of the readers to the volunteer authors, who again cannot be assumed to have specific knowledge.

In summary, the task of writing a paper displays a more complicated structure of dependencies than a prototypical citizen science project. As a result, in the Quench project, we expected to see either additional work done to manage these dependencies, or problems arising from these
dependencies going unmanaged. Identifying the kinds of coordination mechanisms created or needed will be informative for managers of citizen science projects interested in involving volunteers in these additional kinds of scientific work and by extension, to other crowd researchers.

3 Methods

Methodologically, the present study of Zooniverse Quench and Gravity Spy combines collaborative basic research (van de Ven, 2007) and coordination analysis (Crowston and Osborn, 2003). We introduce each in turn.

The present study engaged in collaborative basic research as defined by van de Ven (2007) to understand the design and outcomes of a specific kind of crowdsourcing, online citizen science. We did so through a close collaboration with developers, designers and educators at Zooniverse. Data gathering included questionnaires, interviews and focus groups addressing volunteer motivation and learning as well as trace data analysis on a variety of topics. For Galaxy Zoo Quench in particular, we conducted extensive analysis of the discussion board associated with

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Galaxy Zoo Quench Project Overview</td>
<td><a href="https://quench.galaxyzoo.org/#/project">https://quench.galaxyzoo.org/#/project</a></td>
</tr>
<tr>
<td>b</td>
<td>Major redshift measurement errors in the SDSS stereoscopic pipeline</td>
<td>BGS0000000b/discussions/DGS000021u</td>
</tr>
<tr>
<td>c</td>
<td>Quench Talk Office Hours</td>
<td>BGS000000a/discussions/DGS00001xk</td>
</tr>
<tr>
<td>d</td>
<td>Sample Selection: Post-quenched galaxy and control galaxy</td>
<td>BGS0000001/discussions/DGS00001xy</td>
</tr>
<tr>
<td>e</td>
<td>Classification Result Error</td>
<td>BGS0000008/discussions/DGS000020s</td>
</tr>
<tr>
<td>f</td>
<td>Difference between v5 QS and QC catalogs and their v4 counterparts</td>
<td>BGS0000008/discussions/DGS000022a</td>
</tr>
<tr>
<td>g</td>
<td>Dealing with Sample Selection Issues</td>
<td>BGS0000008/discussions/DGS0000223</td>
</tr>
<tr>
<td>h</td>
<td>Framework and Suggestions for Data Analysis Phase</td>
<td>BGS0000007/discussions/DGS000013u</td>
</tr>
<tr>
<td>i</td>
<td>Temporary Pause in Quench (Resume July 25th)</td>
<td>BGS0000000f/discussions/DGS000023b</td>
</tr>
<tr>
<td>j</td>
<td>Quench project: a proposal aimed at reviving and completing it</td>
<td>BGS0000000e/discussions/DGS000022f</td>
</tr>
</tbody>
</table>

1 Unless a complete URL is given, URLs start https://quenchtalk.galaxyzoo.org/#/boards/
the project to map the history of the project and important events and decisions made over the course of the project. A list of talk posts referenced in the paper is given in Table 1. Additionally, the study draws on several years of prior engagement with the broader citizen science community beyond Zooniverse.

Coordination analysis (Crowston and Osborn, 2003) led us to pay attention to dependencies in the work processes in Galaxy Zoo, Galaxy Zoo Quench and Gravity Spy. The analysis has six steps: defining process boundaries, collecting data, determining actors and resources, determining activities, determining dependencies and model verification. We analyzed our data using this technique, which highlighted dependencies in the system and led to our documentation of the coordination process associated with each project. Equally important, this technique allowed us to specify areas where the management of dependencies broke down, causing coordination problems.

4 Results: Coordination problems in advanced citizen science work

We turn next to an examination of the dependencies, coordination mechanism and observed coordination problems in the two projects. We present first a detailed case study of the Galaxy Zoo Quench project, presenting a synopsis of the history of the project. We then describe the advanced work of Gravity Spy more briefly before turning to a coordination analysis.

4.1 Galaxy Zoo Quench

The Galaxy Zoo Quench project aimed to research, write and publish an academic paper in collaboration with citizen scientists. The topic of the Quench project was “quenched” galaxies, that is, galaxies that have ceased star formation. Galaxies can quench for different reasons and understanding why different kinds of galaxies quench can shed light on the processes of galaxy
evolution. The plan was to code a collection of quenched galaxies for various properties and then compare those galaxies to a matched sample of unquenched galaxies to identify their distinctive properties. Volunteers would classify the galaxies, as in other citizen science projects, conduct data analysis and co-author a professional journal article (Source a, Table 1).

The plan was to complete Phase 1, the classification process, by 1 August 2013 and then proceed to the second phase, data analysis and discussion. The goal for the end of phase 3 was to submit an article to Monthly Notices of the Royal Astronomical Society (MNRAS) Letters, the online portion of the MNRAS Journal, totaling 4-5 pages. Figure 5 presents the flow of data throughout the project, indicating in grey boxes the major outcomes of each of the Phases.

We next describe each of the phases in more detail with attention to the coordination difficulties encountered. Phase 1 consisted of coding galaxies with the characteristics of post-quenched galaxies. The classification included characteristics believed to be related to quench-
ing, specifically galaxies merging, tidal debris, both or neither. Galaxy classification is a mature process that has been used on several citizen scientist projects, mostly notably Galaxy Zoo. Many participants classify each galaxy such that the answers of any one individual has little effect on the outcome. The classification was somewhat delayed, but successfully completed (in that the selected galaxies had classifications) by the end of August 2013 (Source a).

The project as executed included an additional, initially undescribed, phase between the initial coding and data analysis that we have labelled Phase 1b in Figure 5. This phase represents the first collective task in the process, building consensus on the data created in Phase 1 to generate a dataset for analysis in Phase 2. The assumption was that once the galaxies were coded, the results could be used for analysis but in fact it turned out to be a significant undertaking for the group to reach a consensus on the dataset.

First, as the few volunteers who were continuing to Phase 2 started to use the data, they raised concerns about how the final classification was assigned (Sources b, c & d). The initial algorithm used to determine a classification was to take the option selected by the most volunteers, as in other Galaxy Zoo projects (Source e). For example, if “merging” was selected by 3, “tidal debris” by 6, “both” by 2, and “neither” by 7, the galaxy would be classified as “neither”, even though together the other choices that indicate an interesting finding had been chosen more often. This discrepancy was fixed by revising the algorithm to add the count of the three interesting findings together.

Second, the process of revising the dataset led to concerns about the usability of the data. Volunteers were uncertain about the data reliability given the significant changes made between versions. In some cases, errors crept into the files as they were processed by different people. For example, identifiers for the galaxies in the data file are 18-digit numbers. If the file is opened in
Excel (a common tool for citizen scientists since it is widely available), these long numbers could be converted to floating point numbers and truncated, changing the ID, a problem that beset some versions of the data file. There was also inconsistency in variable labeling between datasets, which raised questions about the data provenance (Source f).

A third set of questions arose about the control group of galaxies. To provide a comparison to the quenched galaxies, the scientists involved in the project selected a control group of 3002 galaxies but did so independently from the citizen scientists. The citizen scientists requested clarification on the selection of the control group, which was explained, but doubts remained (Source g). Throughout the project, a recurrent discussion involves the suitability of the sample of galaxies for the study. Sampling had to be done carefully to avoid introducing bias into the results. Participants developed different subsamples based on different selection rules, but did not seem to reach consensus about which sample should be used.

The next phase of the project, Phase 2 in Figure 2, was data analysis. The lead scientist working with the volunteers had encouraged them to “play” with the data and to “have fun and ferret out interesting trends in the data” (Source h). The intent was that the volunteers would explore on their own and then share interesting results with the group, thus experiencing the process of scientific discovery. As noted above, volunteers had already been observed engaging in analyses of other data sets, and the specific volunteers involved seemed capable of such work (i.e., they had the necessary skills for the task of analysis). Furthermore, different analyses could be done in parallel, i.e., there was no dependency between them.

Unexpectedly though, the group encountered difficulties in this phase. Volunteers perceived the task as too open-ended and so did not know how best to proceed. Part of the volunteer feedback on the project was that the project needed more scaffolding of the research process.
Further, during this phase, the lead scientist became unavailable for some time and none of the other scientists on the project could take on a leadership role (Source i). Problems caused by the absence of a single key individual would not be surprising in a conventional team, but it was unexpected in the context of a citizen science project in which members were able—and expected—to make independent contributions.

The volunteers attempted to continue the project, with extensive discussion and various analyses developed. However, the volunteers did not reach a final decision about what should be done, so Phase 2 did not progress to having the desired final set of analyses and a scientific story. As a final analysis was not done, Phase 3, writing, never started.

In 2014, a citizen scientist attempted to revive the project, receiving responses from the other citizen scientists, as well as from 3 scientists (Source j). However, the discussion ended without the project restarting and there were no further posts on the Galaxy Zoo Quench Talk board.

4.2 Gravity Spy advanced work
The main work done in Gravity Spy is classifying glitches into known glitch classes. However, the data set also includes glitches that do not fit a known class, because the glitches in Gravity Spy are evolving as the LIGO detectors are worked on. Some issues are resolved and those classes of glitch disappear from the data, but new kinds of glitches may emerge as the detectors change. Even in the current data, it is believed that there may be additional classes of glitches still waiting to be identified. Accordingly, in the primary labelling, when glitches do not fit one of the twenty-two known glitch classes, volunteers can mark the image subject as “None of the Above” (NoA).
To improve the system to handle these as-yet undescribed classes of glitches and so to better support the LIGO scientists, advanced Gravity Spy volunteers are invited to identify new classes of glitch. They do so by finding and describing sets of glitches with similar appearances that do not fit a pre-existing class. Volunteers can work independently or collaborate with other volunteers in the search. The intent of describing novel glitch classes is that if a new kind of glitch is found to be common, the class can be added as a formal option in the main interface. However, reaching agreement on exemplars, descriptions and names for novel classes poses coordination problems.

One way to document a possible new class of glitch is to develop a collection of examples. Alternately, volunteers may label glitches by posting a comment that includes a hashtag; the system supports searching for glitches by hashtag. However, the grouping into collections or the applied hashtags may not agree, either because the volunteers do not realize that there is already a hashtag in use and so come up with a different one or because they are grouping (and so describing) novel glitches differently.

Volunteers engage in a variety of practices to resolve these differences, such as describing what features of a glitch led them to apply a particular tag or linking a post on a particular glitch to a collection of seemingly similar glitches. Discussion about possible new classes of glitch also take place on separate discussion boards. Volunteers have gone as far as to create structured forms for describing novel glitch classes in an effort to regularize the process.

However, these practices are limited by the design of the Zooniverse system, which was not designed for and so does not directly support volunteer engagement in creating new classes. For example, there is no easy way to let a volunteer know what a particular hashtag is meant to mean or which are active or deprecated. Nor can the applied hashtags be used to determine
which glitches belong to a putative new class, since there is no way to tell when consensus had been reached.

Most importantly, the process that the volunteers go through to develop new classes is only weakly connected to the science team. There is a process by which a potential glitch class can be brought to the attention of the science team, but it is entirely manual. A few of the proposed new classes have been adopted, but most proposals are still pending. As a result, most potential new classes are not coded systematically enough to be of great help to the project scientists as they hunt for the causes of glitches.

5 Discussion

In this section, we interpret the case using coordination theory to identify what kinds of dependencies existed, how those dependencies were managed or not managed and the impact of these dependencies on project performance.

We start with Galaxy Zoo Quench. Phase 1 of this project focused on the tasks of classifying galaxies. Participants could work independently and concurrently to classify the post-quenched galaxies, with minimal dependencies creating constraints on their work. Classification is a mature process, with a sound technological platform and significant history of being completed in Galaxy Zoo, as well as other citizen science projects. The task of looking at an image and clicking on classifications is well-defined. Citizen scientists were both producers and consumers of the data, at least for those continuing to participate beyond Phase 1, so they had significant motivation to complete the task in a timely fashion. As a result, Phase 1 was completed successfully.

Phase 1b was the first collective task in the process. In this phase, the volunteers undertook several tasks to refine the data set for analysis. There is a dependency among these various
data refinements tasks because they are contributing towards a common output. For example, an important part of the analysis was determining which galaxies to include or to exclude in a way that did not introduce biases in the sample that would affect the results. In this phase, the volunteers started to experience difficulties ensuring that the decisions were made consistently. Another interpretation is that the analysis task has a usability dependency with the creation of the data set and the various steps undertaken to refine the data are ways to manage this dependency. However, it was difficult for the volunteers to know what criteria were appropriate, given their lack of expertise in this analysis.

Phase 2 seemed to suffer from more significant coordination problems. First, seemingly to encourage exploration and serendipitous discoveries, the project scientists seem to have provided only general guidance about what analyses should be done, planning to react to the findings of the volunteers. The problem experienced by the volunteers was parallel to the difficulties in developing a suitable data set: even when they have the skills to do an analysis, the volunteers do not have the expertise to know which analyses will be suitable for publication, so they cannot ensure the usability of their output for the next phase, paper writing.

Our initial expectation was that the project would face challenges particularly in Phase 3, due to the complexity of academic writing and level of coordination required to generate a coherent paper. However, as Phase 3 did not start, this case does not provide data to illuminate this question.

In the case of Gravity Spy, the equivalent of Phase 1, classifying glitches into known categories, is progressing smoothly. However, the advanced work of identifying new glitch classes is more problematic. A key problem is that the volunteers propose new classes based on appearance. However, as they do not have a detailed knowledge of the operation of the detectors, they
do not know what those appearances suggest regarding the cause of the glitches or what kinds of problems are more or less serious for the detectors’ operation.

In summary, our analysis of the dependencies in the project suggest a key problem throughout was ensuring the usability of the outputs of each phase of the project for the next phase. In the first phase of the project, the usability of the galaxy classifications was ensured by the design of the coding system and of the Zooniverse system. Even here, issues arose because the coding system was more complicated, requiring a different aggregation technique. Next, creating a data set that was suitable for analysis (Phase 1b) required not only coding galaxies but also selecting a suitable sample, which requires expertise to do in an acceptable way. Finally, in Phase 2 the project ran into unexpected difficulties in finalizing a set of analysis results that would support a paper. Because the volunteers were not experts in astrophysics, it did not seem possible for them to say what analyses would be suitable. Similarly, in Gravity Spy, many of the proposals for new glitch classes are not helpful for the science team.

It is interesting to speculate what would have been the result in the Quench case if the volunteers had been given more specific direction on which analyses to run. However, this approach would have been contrary to the goal of the project, which was to allow volunteers to engage in discovery on their own. In the case of Gravity Spy, the members of the science team do review proposals, but there seems to be no easy way to explain what makes a particular class interesting.

Another way to express the problem experienced is that there was a need to decompose the overall task of developing an analysis into more specific subtasks that different volunteers can work on. However, the volunteers lacked the knowledge of astrophysics (for Quench) or the detectors (for Gravity Spy) needed to do this decomposition and the concomitant recombination.
As a result, when the project scientists are unable to give guidance, the analysis process ground to a halt. The continued interest of the volunteers suggests that the project did not suffer from a lack of motivation on their part. However, the task of managing the usability dependency between analysis and paper writing or detector repairs (and to some extent, between data collection and analysis) turned out not to be one that could be entirely delegated to a volunteer, no matter how motivated.

6 Conclusions

From our initial analysis, we expected that citizen scientists involved in Galaxy Zoo Quench would encounter problems coordinating the work of writing a paper due to the increased coordination demands of this task as compared to the low level of dependencies in typical citizen science work. We did not anticipate problems for Gravity Spy. Unexpectedly, the Quench project encountered significant difficulties at the prior phase of developing a dataset and both projects have difficulties conducting analyses, even though volunteers had an interest, motivation and prior demonstrated ability to conduct analyses and in principle the tasks to be done had low interdependencies as different analyses could be carried out separately.

In the reported case, a key issue throughout is the apparent difficulty for volunteers to assess the usability of their work as a scientific product, a task that requires scientific domain knowledge to be able to perform. In Zooniverse, volunteers thrived when given clear tasks. A few could take on more advanced tasks. However, they were ultimately not able to make decisions about what constituted an interesting dataset or result. Without that input, the project could not progress.

Our analysis leads to several recommendations for how to support advanced work with citizen scientists. Given the reliance of citizen science on volunteers self-selecting tasks, the first
recommendation is that it is necessary to carefully analyze the tasks to ensure that they are feasible for volunteers.

Second, it is important to have a complete accounting of what that tasks are. The analysis done in designing the Quench project seems to have overlooked the work that precedes and surrounds specific analyses. Specifically, the scientists did not seem to account for the work that must be done to ensure that a dataset is usable for analysis or to select which analyses will be interesting to perform. It may be that for experienced researchers, this type of work “goes without saying”, but in a crowd setting, it needed to be spelled out.

A third issue the case highlights is the difference between knowing how to do a task and knowing what users of the output will find useful. It seems that much of the work of ensuring the usability of outputs required tacit knowledge, in this case about what data should look like, what analyses are interesting for publication or what kinds of glitches are problematic. The problem of volunteers evaluating their results has been noted in other crowdsourcing settings (Nagar et al., 2016). For a task to be suitable for crowd work, these evaluation criteria need to be made explicit. Citizen science projects sometimes provide training, which can be quite intensive (e.g., in the details of a data collection protocol or the operation of a detector).

However, it does not seem feasible to train volunteers to develop the kind of insight needed to know what kinds of data or analyses will be interesting for publication. Indeed, even advanced graduate students in a topic can struggle with these questions. Instead, we recommend that projects faced with these sorts of usability dependencies implement feedback mechanisms to quickly evaluate proposals from the volunteers and to provide guidance on improving them, the goal in Gravity Spy and likely the original plan for the Quench project. However, it is hard to
know whether feedback alone would be enough to guide volunteers to a publishable or scientifically-useful analysis result.

And finally, as noted, our analysis of the coordination needed for collaborative writing suggests that the volunteers would have faced significant challenges had they gotten to Phase 3. Exploring the kinds of challenges involved in this sort of work remains a topic for further research.

7 References


