Supporting Crowd Workers: Assembling Resources in Online Citizen Science Projects

Corey Jackson[1], Carsten Østerlund[1], Kevin Crowston[1], Mahboobeh Harandi[1], Gravity Spy Team[2–4]
[1]School of Information Studies, Syracuse University, Syracuse, NY, 13210, USA
[2]Adler Planetarium, Chicago, IL, 60605, USA
[3]Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Dept. of Physics and Astronomy, Northwestern University, 2145 Sheridan Rd, Evanston, IL 60208, USA
[4]Department of Physics, California State University Fullerton, Fullerton, CA, 92831, USA
{cjacks04,costerlu,crowston,mharandi}@syr.edu

ABSTRACT

When individuals join online production communities, they may engage with project resources—e.g., FAQs, tutorials, and comment forums—to learn the practices and norms for contribution. These resources can be large and unorganized, making it difficult for users to know which are relevant. Furthermore, some resources might be more suitable for newcomers while others might work only for experienced ones.

To identify which resources are most relevant for learning, we analyzed the interaction of users with an online citizen science project. Volunteers in this project are occasionally given items with known answers to classify, which allows an estimation of their accuracy on the task. We used this data to determine if resources are used differently by accurate and less accurate users. Methodologically, we applied a Random Forest model to system trace data in order to identify which resources are most predictive of volunteer accuracy. We augmented this analysis with findings from interviews with advanced users.

The resources most predictive of accuracy during early participation seem to center on the social spaces where users gain access to organizational and social practice. In subsequent sessions, predictive activities center on work-related resources that support independent work. This research suggests specific resources might be highlighted to support user development during distinctive stages of a user's history.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

Online communities, situated learning

CSCW'18 TBD

ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

INTRODUCTION

Learning by doing characterizes many online production communities, such as Wikipedia, open source software or citizen science. Even so, these communities typically provide at least some resources to train and socialize new members of the community. Wikipedia, for example, has a set of pages for new editors that provide guidance about how to style Wikipedia articles and that introduce policies and conventions governing participation on the site. Other pages provide members who have more experience with best practices, e.g., for how to interact with newcomers [7]. Q&A communities similarly provide guidance about how to contribute. StackOverflow for example, provides new users with a two-minute tutorial covering how best to formulate questions and the benefits of applying tags. Communities might advertise best practices in frequently asked questions (FAQs) or about pages. In an online citizen science project, Mugar et al. [16] found the comments left by users in the Planet Hunters project served as valuable learning resources for newcomers, as they pointed to specific features of work practice that were lacking in tutorials.

FAQ, how-to pages, and comments may be valuable resources for learning and socialization, but we know little about the process by which members of the community make use of them. Identification of the assemblage of resources (or structured patterns of use) adopted by users helps us see how users as learners make sense of their environment. Knowledge about which resources are useful to users as they learn to contribute could help those who manage online communities know which artifacts to provide or to suggest to users. A complication is that different resources may be useful at different points in a user's interaction. For example, a tutorial could be a valuable learning resource for newcomers during the beginning stages of participation and loses its significance over time.

In this paper, we draw on trace data from an online citizen science project to examine the resources that support participant learning. Specifically, we present an analysis of resource use of more than 850 users in Gravity Spy, a citizen science project hosted on the Zooniverse platform. A distinctive feature of the system is that it collects data on the accuracy of the users on the citizen science task, providing an opportunity to assess how well different users have learned the task. Using this data, we can compare how more or less accurate users assemble learning resources (e.g., tutorials, FAQs, forum posts) at different stages of participation (e.g., novice, sustained, and meta). To our knowledge no other research has examined the behaviours of users at such as fine-grained level.

Specifically, we address the question: *Which project resources seem to best support user learning at different stages of engagement with the project?*

BACKGROUND LITERATURE

Parallel to the increasing use of online learning formats in higher education and the prevalence of crowdsourcing and digital participatory platforms we find a fast-growing literature on e-learning [5, 9, 10, 12, 14, 17]. A large part of this work builds on a practice-based understanding of learning, which does not pin knowledge to the heads of individuals but situates it in a social and material context, whether conceived as communities of practice [13], activity systems [6] or sociomaterial formations [20]. Here, learning is seen as emerging as participants gradually expand their access to activities associated with a specific community. The notion of legitimate peripheral participation connotes this scaffolding of learning practices [13]. In a study of Wikipedia, for instance, Bryant et al. [4] show how novices often start out by simply reading other's articles before they start making their initial contributions. Gradually they gain access more involved tasks.

The literature emerging around online learning formats pays significant attention to the scaffolding of resources. Luckin [14] proposes the notion of a learner centric ecology of resources. Drawing on activity theory, Luckin [14] argues that resources need to be organized and administrated actively if one hopes to make them accessible to the leaner when they need them the most, i.e., in their zone of proximate development. In formal online learning environments, this often involves a curriculum. To characterize informal settings, Siemens [5] introduced a less formalized approach, named "connectivism", focusing on the general connections learners develop and maintain between resources to facilitate continual learning. Here, learning cannot be designed; it can only be "designed for", by creating systems that enable participants to establish and update the connections they develop in an online environment. Extending this idea, network learning [12] focuses on the ways in which information and communication technologies can promote connections between 1) learners, 2) learners and tutors, 3) a learning community and its learning resources. Following this approach should allow participants to gradually assemble an environment promoting their learning.

We notice that Jones's [12] distinction between three types of connections (i.e., between 1) learners, 2) learners and tutors, 3) a learning community and its learning resources) nicely match online course structures in higher education. Learning management systems promote connections between the students and the instructor (e.g., synchronous or asynchronous interactions) and between the students (e.g., discussion boards), as well as providing access to course material (e.g., readings or canned lectures). However, we do not find the same structures on many crowdsourcing and digital participatory platforms. For example, we rarely see a focus on interactions between participants and expert "tutors." In the formal learning environments, resources are most often predefined (e.g., readings, assignments, etc.) organized and maintained by experts in the community.

Previous research on citizen science has shown that information supplied by site (e.g., FAQ, field guides) and those developed by users are valuable learning resources [8, 11, 18]. But, resources generated by the community are also important. Forum posts, for example, educate users about work practice [16]; an essential learning opportunity for users as they become sustained users. Likewise, Mugar et al. [16] showed how citizen scientists learn from accessing traces of other participants' work on the discussion boards. These traces act as proxies for practice, that is, they make visible the socially salient aspects of people's unfolding work practices without requiring the practices themselves to be shared. They help learners get a sense of other people work in situations where they do not have direct access view other's unfolding activities.

In short, the existing literature makes it clear that both stable expert generated resources and evolving user generated resources are important for learning on crowdsourcing and digital participatory platforms. However, we do not have a good sense of which resources will be most effective in supporting participants' learning. Furthermore, which materials are useful may change over time, as resources are not used equally by all users. In comparing two citizen science projects, [8] found users participate differently as they progress, making use of collections and collaborative resources during later stages of their tenure. Newcomers are likely to benefit from a different assemblage of resources compared to more experienced participants. For example, Jackson et al. [11] discovered the forum posts were used mostly by advanced users. This leads us to the question: What assemblages of resources will benefit newcomers? What assemblages of resources will benefit more experienced participants?

THE ZOONIVERSE CITIZEN SCIENCE PLATFORM

Our empirical study is set in the context of an online citizen science project in which users need to learn to perform the citizen science task. Citizen science projects engage members of the public in scientific research [2]. While there are several models of citizen science, the project we investigate here involves volunteers in large-scale scientific data analysis. Such citizen science projects rely on an online worldwide collaboration platform to support the involvement of scientists and the public. The scientists share their research projects with the public who are interested in the science. However, as the volunteers may not have relevant background knowledge, scientists typically also provide multiple learning resources to educate the volunteers. As a result, citizen science projects are often described as informal learning opportunities for contributors [1].

More specifically, we draw on data from Zooniverse. Zooniverse [19] is the largest platform for citizen science projects, hosting more than 70 individuals projects at the time of writing ranging from fields such as astronomy, history, oceanography, and many others. In Zooniverse projects, scientists upload data objects to the platform and ask a series of [scaffolded] questions which collect information about the objects or help filter useful data objects from those which might not be useful for the scientists.

Gravity Spy

In October 2016 Zooniverse, the LIGO Scientific Collaboration (LSC) and other researchers launched Gravity Spy [21], a citizen science project to improve the interferometers used to search for gravitational waves. A challenge for LIGO scientists is high sensitivity of gravity detectors which is needed to search for gravitational waves, but it also records a large quantity of noise (referred to as glitches). The glitches obfuscate or even masquerade as gravitational wave signals, reducing the efficacy of the search. Currently there are more than 20 known classes of glitch with different causes, with the possibility of more classes being identified. Gravity Spy recruits volunteers to classify glitches, which helps to focus the search for their source. The classification interface is shown in Figure 1.

The project makes several advances on the state of the art in citizen science. A particular feature is the implementation of a training regime, in which new users are gradually introduced to the glitch classes: first two, then five, and only after practice, all of the classes. Importantly for our study, as volunteers classify the glitches, they are periodically given gold-standard data to classify (i.e., data with a known classification) in order to assess their accuracy at the classification task and readiness for promotion to a higher level. Another distinctive feature of the system is that advanced users are tasked with looking for new classes of glitches in addition to classifying glitches into existing classes.

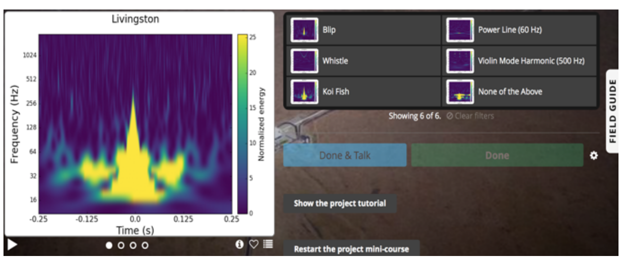


Figure 1. Gravity Spy classification interface. Shown is a screenshot of Level 2; a user can choose one of five glitch classes plus none of the above.

Learning Resources on Gravity Spy. The Gravity Spy project site contains a number of pages through which users can learn about the project, interact and socialize, and organize their work. As examples of the first, the FAQ page (Figure 2) gives users background information about the project such as descriptions of gravitational waves, the LIGO collaboration, and the science behind gravitational wave detection. The field guide (visible to the right of the interface) shows each of the classes of glitch with a description of the key characteristics of each.

As an example of the second, the project hosts five types of discussion boards: science, notes, help, discussion, technical. Science boards (Figure 2, left) contain conversations about the science behind gravitational wave research and related scientific fields. The thread shown in Figure 3 discusses the phenomena of low-frequency bursts observed in one of the interferometers (i.e., Livingston). Notes are conversations about a unique image. When debate exists about what features of an image cause it to be classified as one glitch class versus the other, volunteers leave comments about their reasoning. As an example, in debating whether a glitch was a wandering line or a 1080 lint, one participant states, "Most of the subjects in the collection have a hint of wandering Line mostly above the 1080line. They're quite faint, but if you look closely, you may see the wavy, wandering pattern...." [1]. Help and technical boards are for general questions related to making a contribution or for noting bugs in the interface. The discussion board is a general board for conversations about any subject.

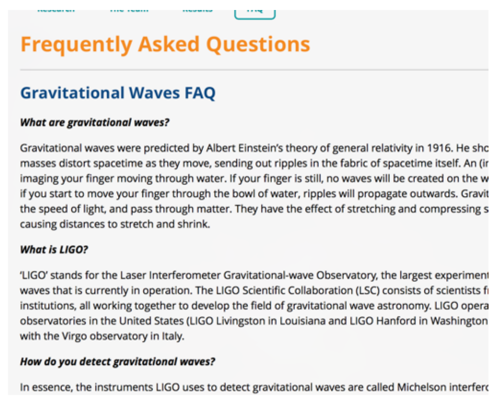
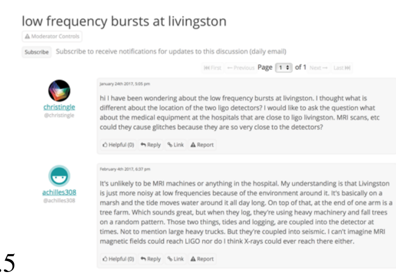


Figure 2. Sample content available on the project FAQ page.



.5

Figure 3. An example comment posted on the science discussion boards.

As an example of the third, the site supports collections for volunteers to keep track of images they find interesting. Collections are also a way for volunteers to organize their independent research activities. As an example, a user created a collection titled "Paired doves timing with alternate morphology" [2] to investigate glitches that share the same 0.4 Hz timing as pairreddoves, but with a different morphology and with a weak amplitude." The collection contains more than ten comments by other users who are also interested in adding to the collection.

DATA AND METHODS

Our primary research question is which project resources support user learning at different stages of engagement with the

[1] https://www.zooniverse.org/projects/zooniverse/gravity-spy/talk/729/207235

[2] https://www.zooniverse.org/projects/zooniverse/gravity-spy/talk/729/160728?comment=277288&page=1

project. We specifically examine resources that support learning to do the citizen science task of classifying glitches. Users may also learn more generally, e.g., about LIGO, about the science of gravitational waves or even about the scientific process, but their immediate engagement with the project is doing classifications, which is a skill that must be learned.

Our study is a mixed-method study involving both quantitative and qualitative data collection and analysis. The quantitative data for this research were collected in January 2017 and cover activities of Gravity Spy users over a four month period beginning in October 2016. The qualitative data consist of interviews with power users conducted over three months.

Data

Quantitative Data

We used two sources of quantitative data—classification data and system interaction data—both drawn from the Gravity Spy system logs. Five months prior to October 2016, the project underwent a beta testing phase where a limited version was available to users. We removed all data from users who contributed during the beta test, since they were already familiar with the project and likely participate differently than the newer cohort of users. The classification dataset contains the classifications users contributed to the project. Included in the dataset are the glitch class chosen by the user (e.g., blip, whistle, etc.), timestamp of the classification, and other metadata about the image such as the image size and glitch type for gold standard images. The system interaction data contains events of users' interaction with pages on the site. When a user clicks on a link to access a new page on the website, an event record is stored. In total, 83 kinds of website events were recorded. The request also contains a timestamp showing the exact day and time the resource was requested. Data were collected anonymously, and include no personally-identifying or demographic data.

Interviews

To augment the quantitative analysis, we conducted interviews with four "power" users of the Gravity Spy project. Interviews were conducted during two periods: the beta stage and in the month following the project's public launch. Interviewees were selected based on their power user status identified by the scientists managing the project. The interviews were semi-structured and lasted approximately one hour. All interviews were transcribed. The purpose of the interviews was to understand how users collaborate. We also asked interviewees to discuss how features such as collections support their work.

Dimension Reduction

To analyze the log data, we first reduced the large number of events (N=83) to a more manageable set. One reason many unique events are recorded is that different links on the site are recorded differently in the database even though they lead to the same page. For example, requesting collections (a resource on the site) from the about menu records the field *type* with the value *about.menu*, while another column in the database shows the name of the page i.e., *collection*. Alternatively, accessing collections from the side bar records the field *type* with the value *side.bar* and a different column in the database records the name of the page, i.e., *collection*. In some cases the field *type* could be *collection*. But the focus of our study is not the path for accessing the resources. We therefore grouped all of these paths to one. We were left with 35 resources, which are described in Table 1.

However, we maintained variations of some resources. For example, we make a distinction between *Collections*, *Collections Collaboration*, and *Collections Personal*. This distinction emerges from our knowledge about the division of labor about collections. The pages contained in *Collections Personal* have to do with a user's individual efforts to maintain a collection while *Collections Collaboration* have to do with collections used by many users. The distinction is important, because maintaining collaborative collections shows a more advanced level of participation in the site e.g., enrolling other members in a work.

Category	Description
About	science team information and gravitational waves summary
Add_*	Comment on the *(Help, Science, or Notes) boards
Blog	Periodic updates about the project
Collections	glitch images users place in a space to be retrieved later
Copy_URL	Copying the URL for a comment post or image
Delete_Comment	Removing a comment from the discussion forums
Edit_Comment	Editing a comment on the discussion forums
Favorites	Accessing the list of images that were favorited
Favorite	Marking an image as a favorite.
Like_Comment	Clicking the "thumbs up" button on a comment
Message_User	Sending a private message to a user
Newsletter	The newsletter for the project where updates are shared
Notifications	Viewing and editing project notifications
Open_Field_Guide	Images and descriptions of the glitch classes
Profile	User's Zooniverse wide user profile.
Report_Post	Mark a post as offensive
Search	Conducting a search
Settings	Customizing user profile information e.g. avatar
Subscribe	Subscribing to a thread
Talk_View	Viewing talk after classifying a glitch
Talk_*Chat	Viewing the talk pages (Help, Notes, Science)
Unfavorite_Image	Unlike an image
View_Image	Open image full-screen
View_Messages	View messages sent to user
View_User	Access the profile of a GS user
View_Metadata	Clicking information associated with an image
Zooniverse_Find	Resources about Zooniverse e.g, publications

Table 1. A list of categories and their descriptions. In total there are thirty-five resources.

Data Aggregation

We aggregated classifications by user into sessions, defined as a set of classification separated by a gap not greater 30 minutes [15]. The intuition is that users tend to come to the system, do one or more classification in a short period, then take a break until later (e.g., the next day). Using session as a unit of analysis allows us to capture the nature of a user's evolving interaction with the system, e.g., by comparing the resources used in the initial session to those used in later sessions. Records of resource use that fell within a session were assigned to the session. To capture use of resources before or after classification, we attributed events up to 25 minutes before or after a session to the nearest session, capturing, for example, if a user visits the FAQ pages before contributing a classification.

As noted above, users are periodically given objects to classify that have known correct answers, called gold standard data. We assess whether the user has learned to do the citizen science task from their accuracy in classifying gold standard data. Specifically, we computed the accuracy of each user in a session as the fraction of user answers that agreed with the known correct answers. Accuracy ranges from 0 to 1. However, not all sessions include gold data, meaning that there are sessions for which we could not compute accuracy. Such sessions are not included in the data set.

We note that the total number of sessions per users is a highly skewed distribution: many users contribute in only one session, while a few contribute in many. Since we were interested in how users' use of resources evolved over their interaction with the system (i.e., from session 1 to later sessions), we removed from the data set users who had only one session. We also removed sessions with fewer than 5 classifications. Finally, since users who contributed in the beta version might experience a different set of features, we removed them from the analysis.

The final data set contains data about 4,530 sessions, recording for each session the user, the session number for the user (i.e., if the session is a user's 1st or nth session), accuracy on gold data within the session and counts of the number of classifications done and of the website events recorded. There are 682 first sessions and 3,848 sessions after a user's first session, representing 832 unique users. There are more users than first sessions because some users did not interact with project features during their first session, but explored them in later sessions. We decided to include these users because their participation intention is the same as a user who had a first session, since they decided to return to the project and to continue to contribute.

Quantitative Data Analysis

Our analysis of the quantitative data was carried out in two phases.

Describing Resource Use

First, we describe the data and report on the statistical features of the dataset, in particular, how user engagement with resources changed during their tenure as contributors to the project. We addressed this question by examining resource use per session number (i.e., which resources are used in users' first sessions, their second, etc.). Because the number of users changes from session to session, we normalized the count of resource use to two measures: *popularity* and *intensity*.

- **popularity:** the number of users who use a resource divided by the total number of users using the system in a particular session
- **intensity:** the number of times a resource was used divided by the number of users who used the resource in a particular session

We then visualize the data to show trends in the use of each resource across sessions.

Identifying Resources Importance for Accuracy

We assess which resources are the most important for modeling the learning process. We would like to understand the difference in resource use between users who have high accuracy versus ones who have low accuracy. A straightforward approach to this problem is to regress accuracy on the use of different resources to identify resources that are significant as predictors of accuracy. However, the relationship between accuracy and resources is non-linear. We therefore applied Random Forest (RF) [3] to identify the most important predictors of accuracy. RF is based on decision trees but is more robust and resistant to overfitting problem. Importantly for our purpose, as it builds a predictive model, the technique produces an importance ranking for the features used to model the data. Our interest lies in these assessments of the importance of the factors. Specifically, we used the randomForest library in R. We report the percent decrease in MSE (%MeanDecreaseAccuracy) for each feature. The %MeanDecreaseAccuracy computes the average decrease of accuracy for each tree when the attribute is removed from the model. Thus, greater %MeanDecreaseAccuracy indicate higher importance to the model.

Qualitative Data Analysis

We present the results from the interviews as a vignette. The emphasis of our qualitative analysis is on describing how power users in Gravity Spy used resources to support their participation and to learn the science of glitch classification. For this paper, two researchers searched through the interview transcripts and identified mentions of project resources. We recorded how the interviewees described the resources, how it supported their participations, and at which point in their tenure with the project the resources were used. We also requested the interviewees' Zooniverse user ids so we could examine the quantitative data about their participation.

FINDINGS

Popularity and Intensity Trends

We first describe how different resources were used over sessions. The most popular resources in the project is *Talk_View* (N = 23,150) followed by the *Open_Field_Guide* (N = 4,177) and *Searching* (N = 3,752). Surprisingly, the *Blog* (N = 1) and *Newsletter* (N = 4) were used on only a handful of occasions in our data set.

In figure 4 we show the aggregated trends in popularity of resource use across fifteen sessions (we limited this number for visualization purposes). Popularity refers to the fraction of users who use a particular resource. As expected, as users learn how participate in the project and to accurately identify glitches, some resources are used less frequently. For example, as users learn the content of the field guide, the need to use that resource decreases. We noticed three interesting points from Figure 4.

First, many of the resources that show a decrease in popularity are learning resources assembled by the science team (i.e., *About*, *Open_Field_Guide*, and *View_Metadata*). The about page, for example teaches users about the project, the science of gravitational waves, and how glitches occur. The field

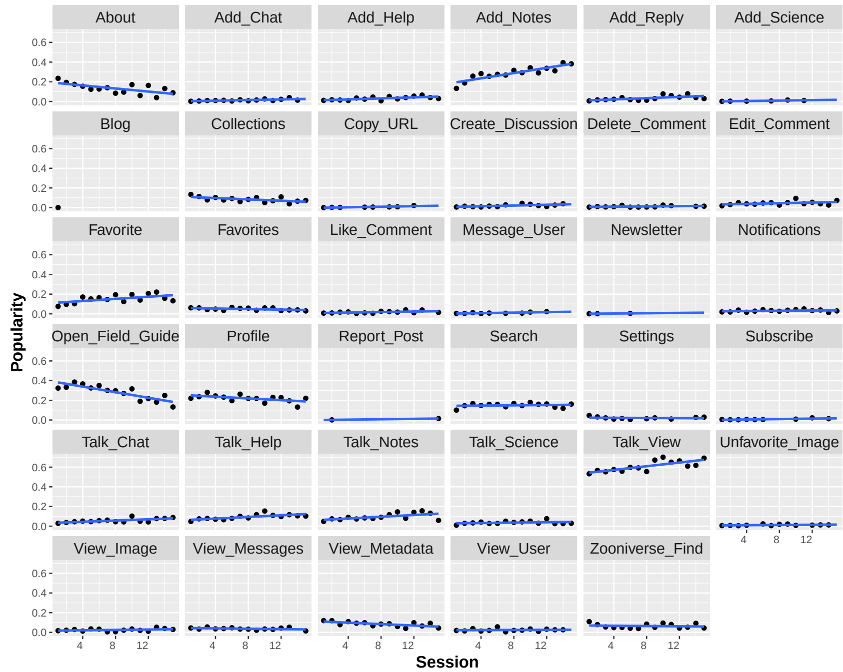


Figure 4. The popularity of resource over time. The x-axis represents sessions and the y-axis represents the popularity or intensity rating for each resource. The blue line shows the trend over time.

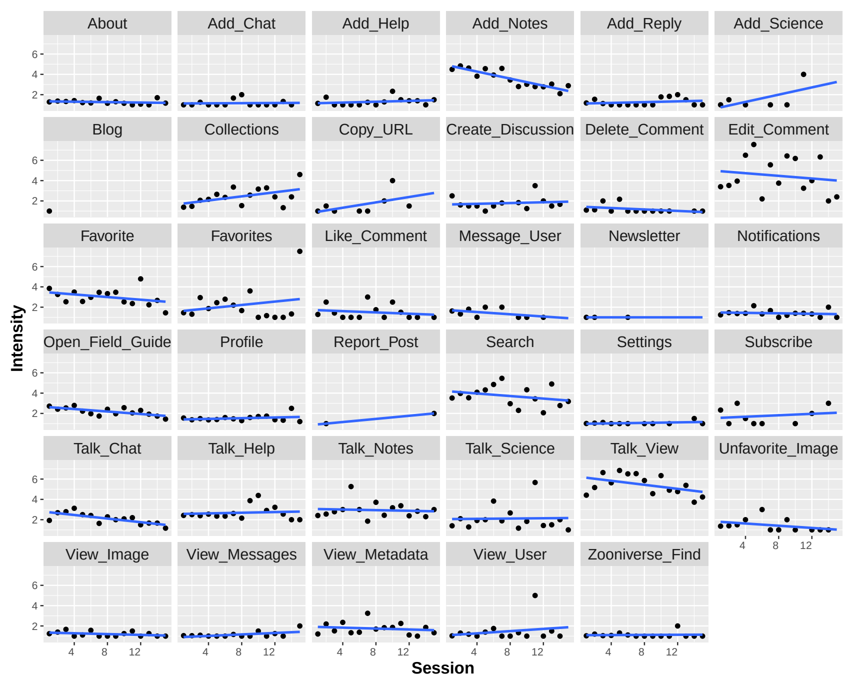


Figure 5. The intensity of resource use over time. The x-axis represents sessions and the y-axis represents the intensity rating for each resource. The blue line shows the trend over time.

guide provides users with a description of the glitch class options that are available as image classifying options. Finally, *View_Metadata* shows information about the image such as related images and the year and month the image was captured.

Second, among the resources increasing in popularity are social spaces where users interact with one another. Increases in the use of *Add_Notes*, for example, indicate a move towards participation that encompasses the social aspects of the site, through which users can ask questions or respond to the comments of other users.

Third, many resources that remain flat over time seem to be those that are not important for learning, e.g., those that have to do with site navigation *View_Image* or social expression *Like_Image*. In both cases, there is low barrier to use and the resources require no scientific knowledge.

Figure 5 shows the intensity of use of each resources over time. Intensity is the number of times per session a resource is used by those who use it. The chart reveals two interesting points.

First, participation in knowledge-related activities. Participation seems to be increasingly purposive on the social spaces of the project. The decrease in *Add_Notes* vs. the increase in *Add_Science* reveal a trend towards knowledge contribution work. The science discussion boards are known to be inaccessible to many users because of scientific jargon. Additionally, the increase in use of collections reveals users curating to identify and work on new glitch classes.

The second observation, unrelated to accuracy or learning, is that we see a increase in resources commonly used to curate new glitch classes (a goal for power users). *Collections* and *Favorites* is interesting because a goal of the project is to have members curate glitches that are likely to be new glitches. *Collections* and *Favorites* are a feature that allows users to examine multiple glitches in a single interface.

Resource Importance with Random Forest

In this section, we describe how we used the Random Forest technique to identify which resource uses were associated with higher accuracy in classification. A first issue was dealing with the distribution of accuracy. Users generally performed well on the gold data, with 92.7% (SD = 13.3) accuracy on average. But the data were highly skewed, as many users had perfect accuracy within a session. We therefore simplified the prediction problem to a binary problem, perfect vs. less than perfect accuracy (1 vs. < 1). This split was chosen because it roughly split the data in half: in the data, 2,149 sessions had perfect accuracy and 2,381 less than perfect.

We created three predictive models using the random forest classifier. The models predict perfect vs. less than perfect accuracy from the 31 web site events plus the count of classifications done. The model is trained on 70% of the data and performance is evaluated on the remaining 30%. The performance of the models are shown in Table 2. Model 1 predicts accuracy for the full dataset, Model 2 just for first sessions, and Model 3, just sessions beyond the first session. Models 2 and 3 are used to compare how resource use changes with experience.

The RF classifier over the test dataset for Model 1 achieved a model accuracy of 75%, correctly predicting 456 perfect and 566 less than perfect sessions. The class accuracy for less than perfect sessions (78%) was slightly better (+7%) than perfect (71%) sessions. According to the RF output, shown in Figure 6, *number classifications* was the most influential variable on the model for improving the quality of the prediction. The next most important attributes were *Open_Field_Guide*,

Observed		*Predicted* 1	< 1	Accuracy
Model 1	1	456	178	71
	< 1	159	566	78
Model 2	1	96	4	96
	< 1	4	101	96
Model 3	1	412	161	72
	< 1	141	441	76

Table 2. Performance of the random forest classifier on three datasets - Model 1 (full dataset), Model 2 (first sessions), and Model 3 (subsequent sessions).

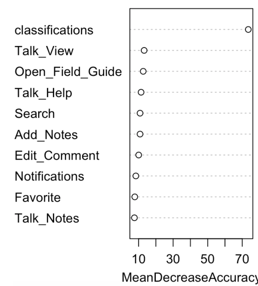


Figure 6. Results of RF with the full dataset.

Talk_View, *Talk_Help*, and *Search*. The first resource is a key guide provided by the science team with descriptions of the glitches. The two talk resources suggest that accurate volunteers also seek information provided by other volunteers. Finally, search can be used to search through the often voluminous talk forums for specific terms, again suggesting an effort to seek resources provided by others.

First vs Subsequent Sessions

The first session dataset that we used to build Model 2 contains 324 sessions with perfect accuracy and 358 with less than perfect accuracy. In the first session the average user contributed 103 (SD = 131.3) classifications. The median was 55 classifications. The accuracy for user classifications in the first session was 96% (SD = 7). The performance of the RF classifier for Model 2 test data was 96% for both classes, correctly identifying all but eight sessions. According to the output of the RF, shown on the left side of Figure 7, *classifications* was most important followed by *Talk_View*, *Search*, *Settings*, and *About*. The first two of these were discussed above. About links to a science team provided description of the project. In contrast to these three, it is difficult to interpret the link from settings to accuracy.

The performance of Model 3 was 73%. Similar to Model 1, Model 3 shows better performance in predicting less than perfect sessions. Again, as shown of the right side of Figure 7, *classifications* dominates the Model 3. The other important attributes are *Open_Field_Guide*, *Talk_View*, *Notifications*, and *Add_Notes*. The first two of these were discussed above. The final two resources indicate a shift to a more social engagement with the project. Add notes indicates the user contribution to a discussion, while notifications allows a user to request to be notified when someone else adds to a discussion.

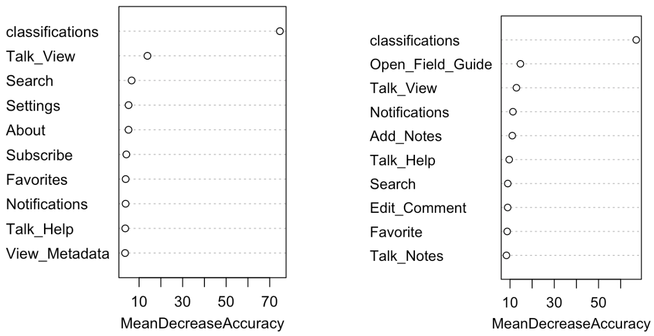


Figure 7. First Session (left) subsequent sessions (right).

Qualitative Findings: The Journey to Power User Status

We next report on findings from the interviews that shed light on the qualitative analysis reported above. Kirsten, Mike, Jason, and Audrey (pseudonyms) are current power users in the Gravity Spy projects. Each has been a contributor since the beta version of the project. Their contribution statistics as of January 1 2017 are shown in Table 3. Their contribution statistics reveal how they have reached power user status, by contributing many classifications, being active beyond the classification task, and achieving high performance over an extended period of time. During the interviews each interviewee mentioned aspects of their participation that helped them become power users and learn how to identify and curate glitches. In addition to the interviews we also visualize the resource use patterns of Jason and Audrey in Figure 8. We noticed two main points about resources that emerged from the interviews and the figures: (1) early participation is spent figuring out the project and the task, and (2) power users consume learning resources early, but eventually generate learning resources for other members.

		Top Activities in Session(s)	
User	**session - classifications - accuracy**	**1**	**2-30**
Audrey	47 - 4,207 - 95%	*Open_Field_Guide*	*Talk_View, Collections, Search, Edit Comments, Talk Chat*
Yoon	135 - 8,609 - 98%	*Talk_View, Edit_Comment, Create_Discussion, Add_Help, Open_Field_Guide*	*Talk_View, Search, Add_Notes, Talk_Notes, Edit_Comment*
Jason	73 - 8,317 - 96%	*Talk_View, Edit_Comment, Open_Field_Guide, Create_Discussion, Add_Help*	*Talk_View, Edit_Comment, Search, Open_Field_Guide, Add_Notes*
Kirsten	82 - 1,639 - 91%	*Add_Notes*	*Add_Notes, Edit_Comment, Search, Talk_View, Add_Reply*

Table 3. Contribution data for the power users we interviewed. Included in the last two columns are activities accessed most frequently in early and subsequent sessions.

In interviews, Kirsten described her early participation and the resources that were valuable as she learned to become a contributor. Kirsten first described the field guide and how the resource provided information about how to identify glitches saying, "when I was first starting out I think I looked at them [*Open_Field_Guide*] pretty frequently just to make sure I was getting it right and understanding what I was looking at." For Kirsten the field guide was an authoritative source from which she could retrieve information to help her performance in identifying glitches. Other interviewees mentioned other resources were also important during the early stages to orient them to the project. Mike for example, noted he consulted journal articles outside the site in addition to information on the site to help hum understand the science behind gravitational wave research. Information found in the *About* pages. Kirsten suggested these resources are important since "...I don't know of anywhere else I can go on say the LIGO collaboration website and find relevant data."

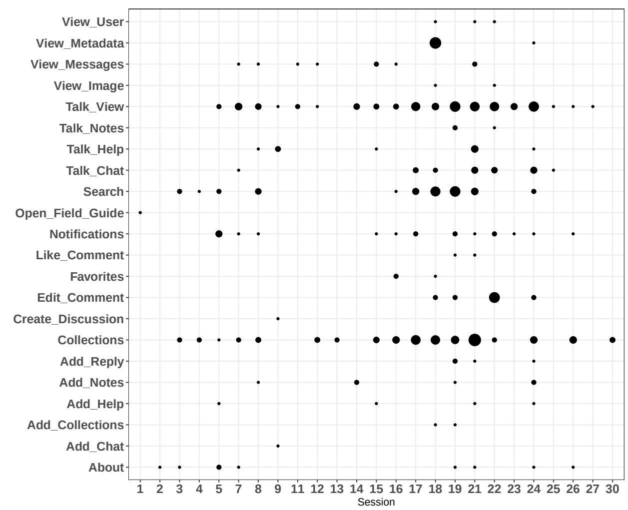


Figure 8. System log activities for Audrey. Each point represents accessing a resources and the size of the point corresponds to the intensity of using the resource. For readability we limit the number of sessions to thirty.

We know a large portion of early participation consist of viewing resources. It is only much later in the project that participants begin to create content. For most interviewees much of the viewing took place when they accessed *Talk_View*. It is a dominant activity for Audrey and Jason, as shown in figures 9 and 8. The *Talk_View* is known to be a place to receive feedback since no information about performance is provided to users. Users can view the work of other contributors or eventually muster the courage to ask questions. Participation in social spaces allows users to improve their accuracy because they can receive feedback from peers about the classifications they submitted Kirsten noted "discussing it with other people can help introduce new understanding of those objects too because maybe there's like a confluence of glitches and they all take place at the same time."

Beyond simply viewing the contributions of others, the roles users choose to adopt in the project imply a different set of features. For example, some users engage in social roles by leaving feedback, answering the questions of newcomers, engaging in conversations about new glitch classes, or curating glitches. Each of these activities imply a different set of features. Curating glitches, for example requires the use of *Collections*, *Favorites*, and *Favorite*. Audrey describes her contribution in curating glitches as an active user of *Collections* and *Search* (verified in Figure 8). Her use of centered on curating new glitch classes, stating, “After I see a certain pattern a few times I create a new collection and search in collections of other volunteers, so that my collection name match with possible others.” Jason became heavily involved in curating glitches as well and described how he wrote his own browsing script (much to the consternation of the Zooniverse software developers) to page through many glitch images . His program goes though the URL’s on the site and he examines them more closely if they contain interesting features. Doing the curation work that Audrey performed, for example, helps sift through the many data objects and provide prototypical glitch examples for other users to consult, Audrey would describe the features of the glitches in the collection, which helps newcomers recognize confusing glitches.

Engaging with other users became important for many power users and they adopted the role of a community manager, welcoming individuals to the project and guiding users to helpful resources. Kirsten’s work centered around providing what she described as feedback on the collections of other users. She described her approach, saying she would leave comments like “hey this is interesting although it doesn’t quite fit, here’s what it kind of looks like but here’s the reason why it doesn’t quite fit I think and then maybe #possiblenewglitch.” This activity explains the many *Add_Notes* in her activity history. Finally, exploring Figures 9 and 8, many of the resources that were important in the early sessions are not found. Audrey and Kirsten for example, didn’t engage beyond the classification work in the first session while Yoon and Jason explored other aspects of the project.

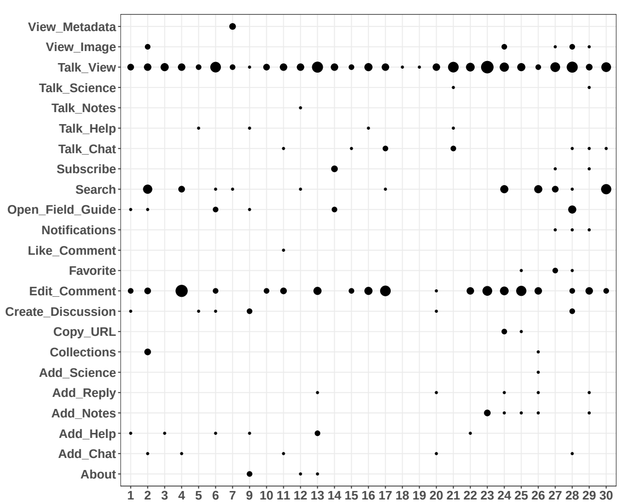


Figure 9. System log activities of Jason. Each point represents accessing a resources and the size of the point corresponds to the intensity of using the resource. For readability we limit the number of sessions to thirty.

DISCUSSION

Assembling Resources for Participation

The three empirical sections reveal how users engage with resources in Gravity Spy. The descriptions of resources use popularity and intensity across sessions revealed that resources constructed by authoritative individuals like the Zoonvierse moderators or the astrophysicists on the science team do seem to help learners increase their knowledge about the search for gravitational waves.

However, the results further revealed that resources that match traditional learning arrangements—materials supplied by individuals with relevant academic backgrounds—lessen in importance over time (e.g., *Open_Field_Guide*). Instead, we see a rise in more involved forms of contribution through social participation or individual organizing.

Social participation takes shape in two ways. First, users engage with others in conversations (*Add_Reply* for example). We can make a distinction between reading (**_View*) and contributing to the social spaces (*Add_**) of the project. Many of the *Add_** spaces increase in popularity and intensity. Second, users begin contributing to spaces where the uninitiated might struggle to comprehend (recall the complexity of discussion in figure 3). In later sessions users engage in spaces of the project dedicated to collaboration (e.g., *Favorites* and *Collection*). Contribution in these spaces reveals intention towards more collaborative and independent work.

The results generated from the random forest algorithm also highlight the movement from individual to social participation. An important finding here is that a significant portion of the variance in accuracy seems to be explained by experience in classifying. One interpretation is simply learning by doing. As well, after users classify gold standard data, they are told if their answers are the same as experts. This continued feedback likely also helps explain the improvement in accuracy.

However, it should be noted that gold standard data are prototypical examples of the classes. These are good for learning but do not prepare users for handling less clear glitches. What to do with these can only be learned through discussions with other users and science team members. Because users can not view other users’ classifications, they must instead rely on what Mugar et al. [16] called practice proxies, that is, discussions on Talk that convey how the task is performed. Specifically, they noted the value of posts on the Planet Hunters fora that point to the shape and form of light curves using context (i.e., what is being observed, e.g., dips in periodicity) and specificity (i.e., where it is observed, e.g., upper right corner).

We find that for more advanced users, the discussion forums are not simply a place to observe practice (*Talk_View*). The importance charts comparing first and subsequent sessions in Figure 7 reveal a shift from observing practice to creating practice. For example, the *Talk_View* page decreases in important in subsequent sessions, but *Add_Notes* becomes important. This shift represents the difference between simply viewing the discussion to adding to it. This interplay supports what [18] describe as different forms of social presence. Here, users relationships to resources change as they gain experience and become more knowledgeable about the task and the community.

Finally, the most interesting observation from the interview results are the activities of power users in adopting roles to sustain the community, which in some respects supports learning. When power users interact on the site, the content they generate serves two purposes, one intentional and the other unintended. As power users curate glitch images, placing them in collections and marking them as favorites, they intentionally aggregate resources to better support their participation in identifying new glitch classes. In the process, they force conversations about the shape and form of glitches, an intended performance of work. An unintended consequence of this work is the trail of resources left for other volunteers to consume. A common issue noted by volunteers is the lack of prototypical examples for the classification task, especially for newly identified classes. As well, there is a lack of guidance that covers marginal examples. As power users collect and debate where these glitches fall, newcomers and users less initiated can learn from these resources.

Supporting Learning

Returning to the background literature, our research contributes to the e-learning literature with a deeper understanding of how user-generated resources contribute to learning in online communities. In Gravity Spy, we see a movement from individual to social and collaborative learning spaces. Recall Jones's [12] discussion of learner-centric ecologies in which there exist connections (1) among learners, (2) between learners and tutors, and (3) between learners and resources we find much of the learning is facilitated though learners and community and its resources. This categorization fits what we see online classes where learners connect to each other through forum posts, learners and experts (tutors or teachers) connect through feedback and learners and community and its resources through resources posted on the forums such as slide decks or notes. In our setting, much of the learning is facilitated by the latter; our results reveal that learning is facilitated primarily through the resources. The other two are less prominent.

A difference in our findings though is that resources produced by authorities seems to be important initially but then decline in use, replaced by access to user-generated content (in our case, talk forums). We think that this situation may apply more generally. On digital participation platforms, participants might encounter a few formal tutorials and readings, but a lot of the attention is placed on user-generated content developed through engagement with the platform (e.g., wiki-articles). For instance, on Wikipedia, users can find resources on how to make their first contribution or how to becomes an administrator, both curated by the community of Wikipedians. Even still, Bryant et al. [4] note that for Wikipedians talk pages are still helpful. These distinctions suggest in the e-learning literature; more attention is given to how learning resources are assembled to support learning, interactions among participations, and relations to the experts.

Implications for Designing Online Communities

Given the empirical results, the question of how best to organize resources on the site is important. Connectivism [5] which suggests learning cannot be designed; it can only be "designed for" by creating infrastructures that allow individuals to make connections to the online environment. Our results highlight movement from the individual to social and collaborative learning resources. These findings suggests a different view on Luckin's [14] learner-centric ecologies. She argues that resources should be actively organized and administered. Our findings suggest that this organization needs to extend to the resources created by users. However, the question of how best to organize user-generated content and to integrate it with the formal training materials for newcomers remains an open question. In this case, the resources become a part of the ecosystem of learning materials but are not easily found. Indeed, the difficulty of navigating the large volume of talk may be the reason for the importance of the search function.

Based on these findings, we suggest that online community managers continually evaluate resources and direct users to new resources that support learning. We also suggest for online production communities, site managers begin to enroll the resources generated by the community by referencing the materials created by users in the authoritative resources such as *About* or *Open_Field_Guide*. This evaluation supports additional research in how users use, create, and assemble resources.

Limitations

As with any research project, there are limitations. The main limitation of our data analysis is that it is based on what the system records in the system logs. The system does not collect the URLs of specific pages, which could have allowed us to draw more fine-grained conclusions about the resources learners use. For example, analysis might reveal more specifically which *Talk_View* pages are important.

As well, we only examined the resources which are available or generated in the Gravity Spy project. We know from interviews that some users consult resources external to project site, e.g., scientific publications and videos, which might help users to have a better understanding of gravitational waves and glitch detection. However, we have no way to track the use of such resources.

Additionally, we removed sessions where users had no gold data. These are likely sessions in which users had a small number of classifications, nevertheless, the activities of these users could be important in efforts to draw conclusions about the resource used to support learning in Gravity Spy.

ACKNOWLEDGEMENTS

Without the Zooniverse volunteers who worked on the projects, there would be no paper. Many thanks to the Zooniverse team for access to data. This material is based on work supported by the National Science Foundation under Grant No. IIS xx-xxxxx.

REFERENCES

1. R Bonney, H Ballard, R Jordan, E McCallie, T Phillips, J Shirk, and C C Wilderman. 2009a. *Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education*. Technical Report.

2. Rick Bonney, Caren B Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V Rosenberg, and Jennifer Shirk. 2009b. Citizen Science: A Developing

Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience* 59, 11 (Dec. 2009), 977–984.

3. Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (2001), 5–32.

4. Susan L. Bryant, Andrea Forte, and Amy Bruckman. 2005. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work (GROUP '05)*. ACM, New York, NY, USA, 1–10. DOI: **http://dx.doi.org/10.1145/1099203.1099205**

5. Stephen Downes. 2005. E-learning 2.0. *Elearn magazine* 2005, 10 (2005), 1.

6. Y Engeström. 1987. *Learning by expanding*. Cambridge University Press.

7. Aaron Halfaker, Aniket Kittur, and John Riedl. 2011. Don'T Bite the Newbies: How Reverts Affect the Quantity and Quality of Wikipedia Work. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*. ACM, New York, NY, USA, 163–172. DOI: **http://dx.doi.org/10.1145/2038558.2038585**

8. Katie DeVries Hassman, Gabriel Mugar, Carsten Østerlund, and Corey Jackson. 2013. Learning at the Seafloor, Looking at the Sky: The Relationship Between Individual Tasks and Collaborative Engagement in Two Citizen Science Projects. *In proceedings for 10th International Conference on Computer Supported Collaborative Learning* (June 2013).

9. C Haythornthwaite. 2014. New Media, New Literacies, and New Forms of Learning. *International Journal of Learning and Media* 4, 3-4 (2014), 1–8.

10. Caroline A Haythornthwaite and Richard 1953 April 1 Andrews. 2011. *E-learning Theory & Practice*. Los Angeles : Sage.

11. Corey Brian Jackson, Carsten Østerlund, Gabriel Mugar, Katie DeVries Hassman, and Kevin Crowston. 2014. Motivations for Sustained Participation in Crowdsourcing: Case Studies of Citizen Science on the Role of Talk. In *2015 48th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 1624–1634.

12. C Jones and M de Laat. 2016. Network Learning. In *The SAGE Handbook of E-learning Research*, C Haythornthwaite, R Andrews, J Fransman, and E M Meyers (Eds.). 44–61.

13. Jean Lave and Etienne Wenger. 1991. *Situated Learning. Legitimate Peripheral Participation*. Cambridge University Press., NY.

14. Rosemary Luckin. 2008. The learner centric ecology of resources: A framework for using technology to scaffold learning. *Computers & Education* 50, 2 (Feb. 2008), 449–462.

15. Andrew Mao, Ece Kamar, and Eric Horvitz. 2013. Why Stop Now? Predicting Worker Engagement in Online Crowdsourcing. In *Proceedings of the First AAAI Conference on Human Computation and Crowdsourcing, HCOMP 2013, November 7-9, 2013, Palm Springs, CA, USA*.

16. Gabriel Mugar, Carsten Østerlund, Katie DeVries Hassman, Kevin Crowston, and Corey Brian Jackson. 2014. Planet Hunters and Seafloor Explorers: Legitimate Peripheral Participation Through Practice Proxies in Online Citizen Science. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 109–119. DOI: **http://dx.doi.org/10.1145/2531602.2531721**

17. Carsten Østerlund and Paul Carlile. 2005. Relations in Practice: Sorting Through Practice Theories on Knowledge Sharing in Complex Organizations. *The Information Society* 21, 2 (April 2005), 91–107.

18. Carsten Østerlund, Gabriel Mugar, Corey Brian Jackson, Katie DeVries Hassman, and Kevin Crowston. 2014. Socializing the Crowd: Learning to talk in citizen science. In *Academy of Management Annual Meeting, OCIS Division*. Philadelphia, PA.

19. Robert Simpson, Kevin R. Page, and David De Roure. 2014. Zooniverse: Observing the World's Largest Citizen Science Platform. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion)*. ACM, New York, NY, USA, 1049–1054. DOI: **http://dx.doi.org/10.1145/2567948.2579215**

20. Estrid Sørensen. 2007. The Time of Materiality. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 8, 1 (2007). **http://www.qualitative-research.net/index.php/fqs/article/view/207**

21. Michael Zevin, Scott Coughlin, Sara Bahaadini, Emre Besler, Neda Rohani, Sarah Allen, Miriam Cabero, Kevin Crowston, Aggelos Katsaggelos, Shane Larson, Tae Kyoung Lee, Chris Lintott, Tyson Littenberg, Andrew Lundgren, Carsten Øesterlund, Joshua Smith, Laura Trouille, and Vicky Kalogera. 2016. Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science. *arXiv. gr-qc* (Oct. 2016), 1–27.