How is success defined and measured in online citizen science? A case study of Zooniverse projects

Joe Cox, Eun Young Oh, Brooke Simmons, Chris Lintott,

Karen Masters, Gary Graham, Anita Greenhill and Kate Holmes

Abstract

While the literature highlights a wide variety of potential citizen science project outcomes, no prior studies have systematically assessed performance against a comprehensive set of criteria. Our study is the first to propose a novel framework for assessing citizen science projects against multiple dimensions of success. We apply this framework to a sample of projects forming part of the online Zooniverse platform and position these projects against a ‘success matrix’ measuring both contribution to science and public engagement levels relative to others in the sample. Our results indicate that better performing projects tend to be those which are more established, as well as those in the area of astronomy. Implications for citizen science practitioners include the need to consider the impact of core competencies on project performance, as well as the importance of relationships between the central organisation and science teams.

Keywords: Citizen science; Online; Project outcomes; Success factors; Zooniverse

1: Introduction

The last decade has seen a number of significant developments and innovations in online citizen science, not least of which being the creation of the Zooniverse; a cluster of projects that source volunteer contributors to analyse and interpret large datasets. Although these data are too complex to interpret using computer algorithms [1], the analytical tasks volunteers are asked to complete are sufficiently simple that members of the public can engage meaningfully without having a specialist knowledge or background in science [2]. The Zooniverse platform which exists today grew out of the original Galaxy Zoo project, which has been identified by a number of scientific organisations around the world as being 'high profile' [3], 'well-known' [4] and 'successful' [5]. The Galaxy Zoo project received 70,000 classifications per hour within 24 hours of its initial launch and more than 50,000,000 classifications within its first year. As a consequence of the popularity of this initial project, the Zooniverse platform[1] has subsequently launched an increasingly diverse range of other projects and now has more than 1.1 million registered volunteers from around the world.

Zooniverse projects are united by two distinct aims and objectives, the first of which being to solve specific scientific problems by serving as a reduction tool for data (and labour) intensive science and transforming raw user inputs into a 'data product' for use in research [6]. This is achieved by making 'the best use of knowledge and skills of volunteers rather than their computers', while also benefiting from the serendipitous discoveries that often emerge from the visual inspection of datasets [7]. The second core objective of Zooniverse projects is to engage with the public in order to educate and change attitudes towards science. This goal manifests itself in practice through the use of blogs, Twitter feeds and other social media outlets, as well as outreach and education programmes such as 'Zooteach[2]'.

A number of prior studies of citizen science undertaken by Wiggins & Crowston [8; 9; 10] have modelled the organisational structures of projects and created a typology of activates based around variations in goals and tasks. This work highlights the significant heterogeneity that exists between online citizen science projects, which often limits the extent to which one can be directly compared against another. The aim of this paper is to at least partly address the lack of common criteria that can be

[1] http://www.zooniverse.org
[2] http://teach.zooniverse.org

used to compare and contrast the performance of online citizen science projects within the diverse online ecosystem of the Zooniverse. We therefore set the following specific research questions:

(i) How can measures of success and outputs from a citizen science project be defined?

(ii) What is the relative positioning of Zooniverse projects against these measures of success?

Our study is novel in this approach for a number of reasons. First, we define a representative set of project-level outcomes highlighted by citizen science literature combined with a number of unique measures to assess the extent to which a project has been successful. Second, we use these measures to specify a unique citizen science 'performance matrix' and use this to assess the relative performance of a sample of 17 Zooniverse projects[3] spanning a range of activities and scientific disciplines. By developing an understanding of the differences between better and less well performing projects, the work presented in this paper will be of value to citizen science practitioners in identifying and learning from cases of 'best practice' in the field.

2. Literature Review

Although the literature on citizen science generally acknowledges the broad outcomes of scientific contribution and public engagement, a number of authors have interpreted these outcomes in different ways. With respect to scientific contribution, the quality, size and/or completeness of data generated is frequently mentioned as a key project outcome [11, 12, 13, 14]. Although assessing impact through publication and citation counts can potentially be problematic [15] and may reflect other project-specific factors, publications and citations in peer-reviewed academic journals can be argued to represent an objective measure of the scientific value of the data generated by the project [16]. Indeed, the co-authorship of academic papers is a means by which well-functioning citizen science platforms can formally recognise the participation of volunteers [17] and incentivise more valuable contributions [9]. Additionally, effective *project design and resource allocation* is highlighted by a number of other authors as an important aspect of producing high quality data output, i.e. the extent to which projects are intuitive to use, break down large tasks to an appropriately small scale and successfully match teams of

[3] The specific projects considered are Galaxy Zoo 1-4, Moon Zoo, Planet 4, Planet Hunters, Solar Stormwatch, The Milky Way Project, Bat Detective, Seafloor Explorer, Snapshot Serengeti, Whale FM, Ancient Lives, Cyclone Center, Cell Slider, Old Weather (plus the Andromeda Project).

scientists and participants to sets of tasks [18]. Raddick *et al.* [17] at least partly defines successful projects in terms of the calibration of user contributions, i.e. the extent to which appropriately sophisticated algorithms are employed to convert the raw data provided by participants into meaningful scientific insight [9]. Other measures of effective project design and resource allocation include the provision of adequate training [19], the division of effort between volunteers [20] and the extent to which accurate data can be collected at a lower cost [21].

With respect to the second broad aim of public engagement, several authors highlight the importance of *dissemination and feedback* as a key project outcome. This relates to informing participants about the ways in which their data have been used [18], while also serving as a means by which volunteers can be rewarded for their participation [1]. Bauer & Jensen [22] also highlight the importance of organised public outreach events to achieve these objectives. A parallel strand of public engagement is the extent to which citizen science projects lead to greater levels of *participation and opportunities for learning*. Participation can be measured in a number of different ways, such as the extent to which a project succeeds in generating a critical mass of volunteers [9; 23] or by a project's ability to sustain levels of engagement over longer periods [20]. While providing opportunities to enhance understanding of science is widely identified as a key outcome of citizen science projects [24; 25; 26], changes in scientific literacy are often extremely difficult to measure in practice. This is because direct measures, such as enhanced understanding of science content and processes [17] cannot be determined without extensive longitudinal research conducted with volunteers themselves. However, a number of authors have suggested that effective proxies for scientific literacy are measures of participation such as the duration of involvement in projects [16] and/or through observing changing patterns of communication, feedback and participation in public forums [27].

3. Data

Our comparative analysis of Zooniverse projects is based on the implementation of a 'positioning matrix' to identify better and less well performing projects, as well as the key differences between them. The two main dimensions we use to position projects on this matrix are *contribution to science* and *public engagement*. The suggested sub-criteria making up the higher level elements of the matrix are presented in Table 1, along with a range of suggested measurements and/or proxies of project performance against these sub-criteria. Our measures are derived from raw classification files and project backups generated by the Zooniverse platform, as well as web analytics for individual projects, blogs, and Twitter feeds. Data for retired or inactive projects encompass all active project dates; data for ongoing projects were collected between the 21st September and the 2nd October 2014. In each case, we use the term ‘subject’ to refer to a single data artefact, such as an image or an audio clip, while the term ‘classification’ refers to the completion of a single unit of analysis by a volunteer. Community engagement measures for these projects are calculated by assigning a relevant unit of engagement (e.g., a forum post, a new Twitter follower, or a blog view) on the condition that the project that was active on the date the new engagement was registered. For projects with periods of inactivity, statistics such as blog views are only counted if they fall within the active period of the project.

For many measures of project outcomes, we report rates of activity over time as opposed to raw numbers; both in terms of the active project duration (the length of time that the project has actively accepted new classifications) and project age (the length of time between the start of the project and October 2014, which may include periods of inactivity). These are used as appropriate depending on whether the particular performance measure can only occur whilst the project is active (e.g. classification activity) or after the project has finished accepting new classifications (e.g. publications). In order to account for nonlinearity in the growth of these activities over time, we calculate activity rates on the basis of dividing raw figures by active project duration or age squared. This simple measure allows us to broadly account for the expected rise in publication rates for scientific projects over time observed in other studies [28; 29] and at least partly remove the bias caused by directly comparing longer running projects against projects with shorter durations.

The criterion for inclusion in the sample is that projects should have been launched at least eighteen months prior to this study. Although time may not be the only constraint upon publication activity, we consider this to be the minimum project age that would allow a chance for science teams to publish at least some output given the median observed period of 21 months for project science teams to publish their first paper. According to this criterion, a range of data were sampled from 17 online citizen science projects with an active period of at least one year, plus one additional project with a duration of approximately one month (The Andromeda Project). The Andromeda Project is exceptional amongst projects in our study due to its short duration, which makes it difficult to compare against other projects due to the unusually high level of public engagement received over such an abnormally short space of time. While the enormous success of the Andromeda Project should clearly be acknowledged, it remains a significant (positive) outlier in this sample group for a number of reasons. We thus exclude the project from the full analysis and limit our direct comparison to those projects with an active period of greater than 1 year. The scores awarded to each of the 17 remaining projects are calculated by comparing projects relative to the leading performer, meaning that at least one project always receives a score of '1' against each measure. Projects are broken down into four broad categories, namely (A) Galaxy Zoo; (B) Other Astronomy; (C) Ecology and (D) Other.

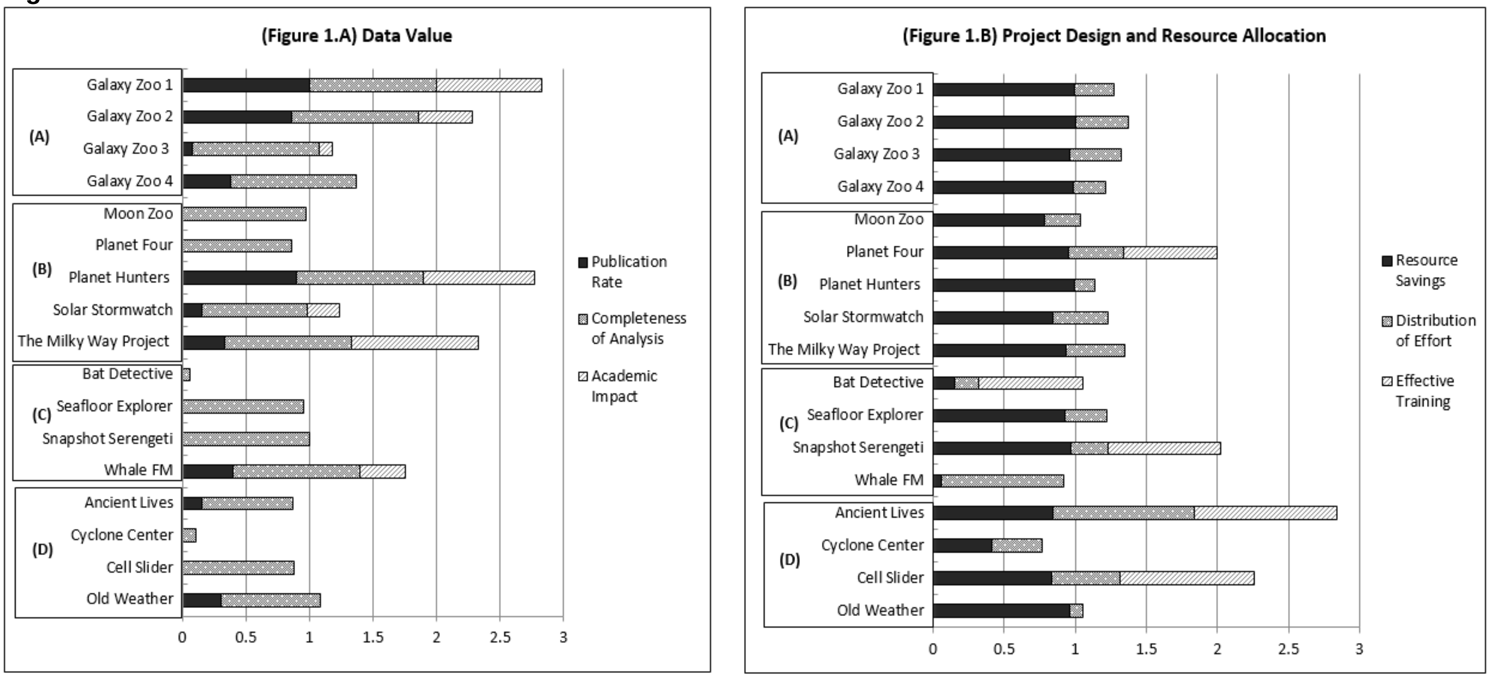
Table 1: Elements of citizen science success matrix

Matrix Element	Performance Indicator	Citations	Measurement	Proxy	Description
Contribution to Science	Data Value	Bonney *et al.* (2009) Cashnman *et al.* (2008) Cohn (2008) Dai *et al.* (2010) Gardiner (2012) Raddick *et al.* (2009) Riesch & Potter (2014) Sheppard & Terveen (2011) Silvertown (2009) Wiggins & Crowston (2011)	Publication Rate	$\frac{Number\ of\ published\ papers}{(Project\ age)^2}$	Total number of papers published divided by the square of project age. In fields where peer-reviewed journal articles are the norm, this includes only published or in-press peer-reviewed articles.
			Completeness of Analysis	$\frac{Number\ of\ classifications}{Target\ number\ of\ classifications}$	Total number of classifications received by the project divided by the target number of classifications per subject. The target is determined as the number of classifications per subject required to achieve an acceptable level of scientific and statistical validity.
			Academic Impact	$\frac{Number\ of\ citations\ per\ publication}{(Project\ age)^2}$	Total number of citations received per publication divided by the square of project age.
	Project Design and Resource Allocation	Dai *et al.* (2010) Franzoni & Sauermann (2014) Raddick *et al.* (2009) Rotman *et al.* (2012) Wiggins & Crowston (2011)	Resource Savings	$1 - \left(\frac{\text{Active project duration}}{\text{One person workload}}\right)$	Active project duration divided by the number of *weeks* that a professional would need to work as *a full time (35 hours per week)* to complete all classifications recorded on the project.
			Distribution of Effort	$1 - (Gini\ coefficient)$	Measures equality in the distribution of classifications across users.
			Effective Training	$1 - \left(\frac{\text{Volunteers who only complete tutorial}}{\text{Total number of volunteers}}\right)$	The proportion of volunteers who go on to complete at least once classification after completing the tutorial. Note that we do not report data for some projects due to the absence of a tutorial or lack of reliable data on completion rates.
Public Engagement	Dissemination and Feedback	Bauer & Jensen (2011) Elam & Bertilsson (2003) Franzoni & Sauermann (2014) Powell & Colin (2008) Rotman *et al.* (2012) Silvertown (2009) Wiggins & Crowston (2010) Wiggins & Crowston (2011)	Collaboration	$\frac{Number\ of\ papers\ with\ citizen\ scientist\ coauthors}{(Project\ age)^2}$	Total number of papers where the list of authors contains at least one citizen scientist author divided by project age squared.
			Communication	$\frac{Number\ of\ project\ Tweets + blog\ posts + Talk\ posts}{(Project\ active\ period)^2}$	Sum total of project communication activity measured across multiple channels divided by project active period squared.
			Interaction	$\frac{Number\ of\ science\ team\ Talk\ posts + blog\ replies}{(Project\ active\ period)^2}$	Sum total of occurrences of interaction between the science team and volunteers divided by project active period squared.
	Participation and Opportunities for Learning	Bonney *et al.* (2009) Brossard (2005) Cronge *et al.* (2011) Phillips *et al.* (2014) Raddick *et al.* (2009) Trumbel (2000) Wiggins & Crowston (2010)	Project Appeal	$\frac{Number\ of\ volunteers}{(Project\ active\ period)^2}$	Total number of volunteers who have contributed to the project divided by project active period squared.
			Sustained Engagement	$\frac{Median\ volunteer\ active\ period}{(Project\ active\ period)^2}$	Median time interval (in weeks) between a registered user's first and last recorded classification divided by project active period squared.
			Public Contribution	$\frac{Median\ classifications\ per\ volunteer}{(Project\ active\ period)^2}$	Median number of classifications per registered volunteer divided by project active period squared.

4. Analysis

Figure 1 contains project-level data on *contribution to science*. Performance against the *data value* indicator is presented in Figure 1.A and measures the extent to which the output of the various projects has contributed to the stock of science knowledge in their respective fields. The data show that almost half the projects in the sample (7/17) have not produced any publications to date. As a result, these projects receive a score of zero for both *publication rate* and the *academic impact*, meaning that performance is somewhat unevenly distributed against this measure within our sample. Projects that have scored well here are mainly those related to astronomy, especially the early Galaxy Zoo projects, the Milky Way project and Planet Hunters, while notable exceptions from outside astronomy are Old Weather and Whale FM. The three astronomy projects explicitly mentioned here represent the 'early' Zooniverse projects, suggesting the strong performance is at least partially an effect of time rather than just subject area. This is possibly driven by variations in publication rates across scientific fields documented elsewhere in the literature [30; 31]. By comparison, the *completeness of analysis* is considerably more evenly distributed across projects, although two significant outliers are Bat Detective and Cyclone Center, which have both received relatively low numbers of classifications per subject relative to their target. Generally, it is clear that other non-astronomy projects tend to score somewhat unfavourably against these measures compared to astronomy projects.

Figure 1: Contribution to Science



Note: (A) Galaxy Zoo; (B) Other Astronomy; (C) Ecology; (D) Other

Figure 1.B reports data on *project design and allocation of resources*. The measures in this indicator are intended to act as a proxy for the extent to which effective project design and organisation allows volunteer input to achieve maximum impact. It is clear that all Zooniverse projects perform roughly equally on two counts, namely the (in)equitable distribution of volunteer effort and cost savings, measured in terms of the amount of time it would otherwise have taken a professional scientist to have analysed the same quantity of data. There are no clear patterns in the differences observed between projects across subject areas and of different durations, although it should be noted that the projects associated with the lowest *resource savings* are the only two audio-based projects in the sample (Bat Detective and Whale FM), which both have received relatively low numbers of classifications. The inequitable distribution of volunteer effort is highlighted by relatively low levels of equality for most projects (mean value of [1- Gini coefficient] = 0.19), which indicates that the distribution of effort across users is long-tailed. Notable exceptions with more equitable distributions of effort are Ancient Lives and Whale FM. For both of these projects, the number of active hours and classifications per user are relatively low, suggesting these projects have a high incidence of users leaving these projects after supplying a low number of classifications

Most Zooniverse projects are broadly similar in the extent to which they lead to *cost savings*, with an average across projects of around 34 full-time working years saved due to the involvement of volunteers[4]. Even these figures are likely to understate the value of data analysis by large numbers of contributors given the potential for unexpected discoveries and opportunities for education and public engagement. Data on training provision are not available in all cases, either due to the lack of a tutorial feature for a given project or because tutorial participation was not recorded. For those projects where data are available, it is clear that a relatively high proportion of users completing training exercises go on to provide full classifications. Projects such as Ancient Lives, Cell Slider, Snapshot Serengeti and Bat Detective outperform others according to this measure, which may be indicative of particularly effective design in the tutorials of these projects. In practice, it is likely that differences in the tutorial completion and large numbers of missing values are a result of changes in the way tutorials are designed and classifications recorded over time; where both practices have only been standardised across Zooniverse projects relatively recently.

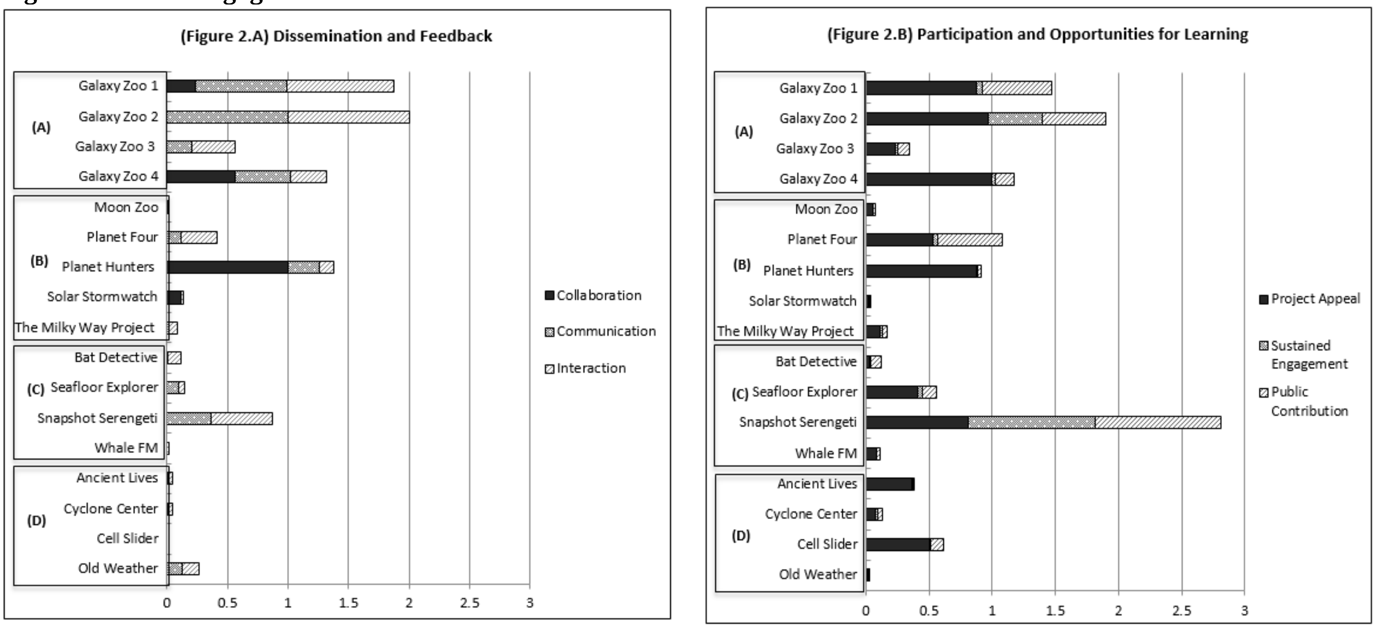
Figure 2 contains data relating to the *public engagement* element of our success matrix, with Figure 2.A reporting project performance against *dissemination and feedback*. An aggregate measure of activity on blogs, Twitter and Talk pages is used as a proxy for *communication*, while an aggregated measure of blog replies and Talk posts made by members of the science teams is used as a proxy for *interaction*. We choose to construct a composite indicator of activity in this way in order to at least partly counter any bias caused by deliberate de-prioritisation of certain individual channels as part of each project's engagement strategy, as well as differences in the culture of social media use between scientific disciplines [32]. Projects that are successful in terms of *communication* and *interaction* are predominantly those in the area of astronomy, especially Galaxy Zoo where activity levels are consistently high across all three channels. Outside of Galaxy Zoo, only Snapshot Serengeti scores well against these measures, largely because of relatively high levels of blogging compared with other projects. Comparatively lower levels of activity are observed for nature and humanities projects. This is especially true for Whale FM, Ancient Lives and Cyclone Center, which receive lower scores mainly due to low numbers of Tweets and forum posts. A potential reason for these differences is variation in Talk activity, both in terms of use by volunteers and science teams. The implementation of Talk is very different between Zooniverse projects: some explicitly require a decision on whether to discuss subjects

[4] The average length of time it would take a professional to classify the same amount of data that has been recorded against a project in our sample is approximately 37 years, whereas the average project active period is 2.4 years.

or provide an obvious link to do so, while others do not. It should also be noted that the Cell Slider project scores zero against these criteria due to an absence of a project blog, Twitter account or Talk feature. This was a conscious decision driven by a concern over the discussion of medical issues without proper moderation, although communication and interaction relating to this project is likely to occur via other channels outside of Zooniverse control.

The final and somewhat less frequently observed outcome is *collaboration*, which we measure by aggregating the number of papers that have been published listing citizen science contributors as co-authors. This measures instances where project volunteers who have either progressed or been invited to participate in more advanced work with professional scientists and is observed only for astronomy related projects; specifically variants of Galaxy Zoo, Planet Hunters and Solar Stormwatch. Although this could be argued to represent a fairly high bar for success and often occurs only as a result of particularly significant and unusual discoveries, this measure can nonetheless be argued to be at least partially related to the richness of the project data set and the engagement level of volunteers.

Figure 2: Public Engagement

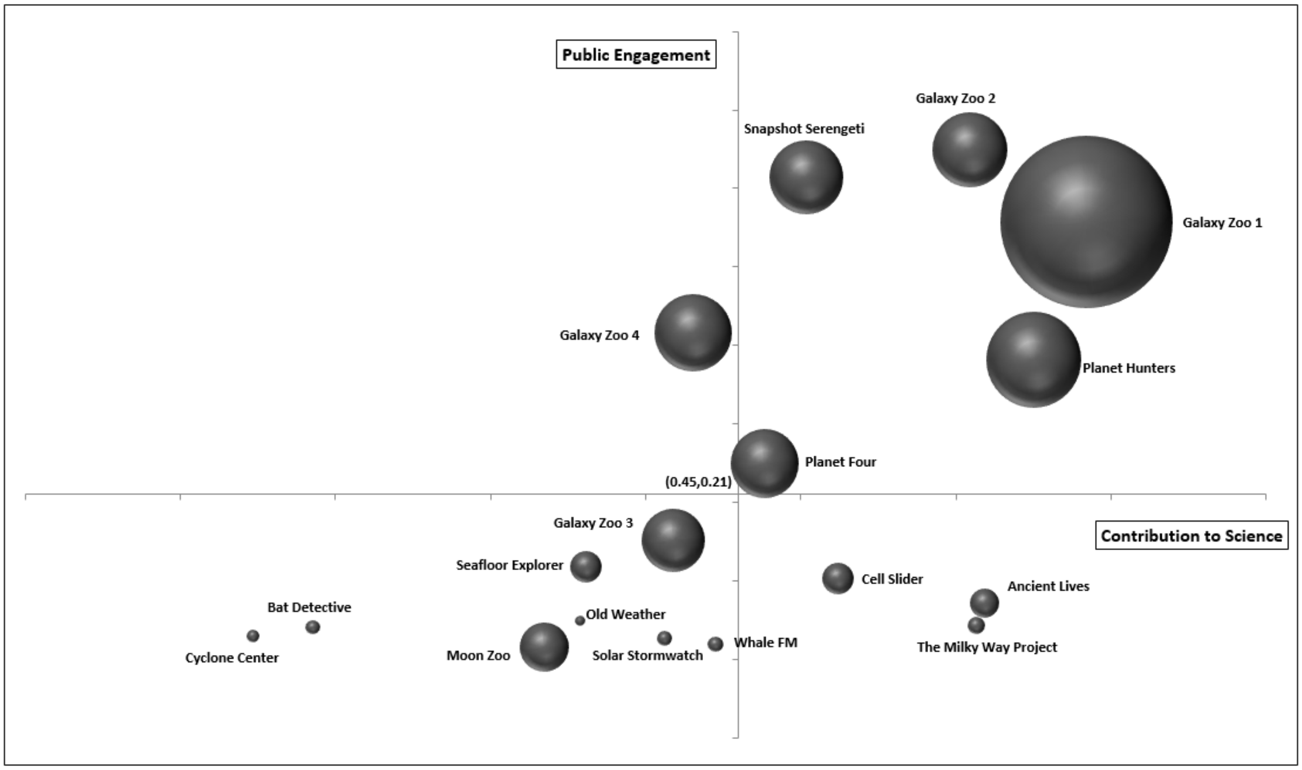


Note: (A) Galaxy Zoo; (B) Other Astronomy; (C) Ecology; (D) Other

Figure 2.B contains data relating to our second measure of public engagement, which is *participation and opportunities for learning*. The first dimension of this is the sustained engagement of volunteers, measured as the median duration between the first and last classifications received by contributors. Performance against this measure is dominated by Snapshot Serengeti, which has a median number of 4.3 hours of sustained engagement per volunteer versus an average of just over 30 minutes for all other projects. Snapshot Serengeti once again dominates the *public contribution* measure, with a median of 61 classifications provided by each volunteer over a comparatively short active period compared to a median of around 21 classifications per user on average for other projects. A potential reason for this variation may be due to the different lengths of time it takes to complete a single classification. *Project appeal*, measured by the total number of contributors to the project standardised by active period, once again indicates a strong performance for Snapshot Serengeti, although comparable performances are observed for most Galaxy Zoo projects and Planet Hunters. Overall, these measures show a significant contrast between projects that have strong project appeal and those that do not.

Figure 3 contains the success matrix reflecting aggregated performance against *contribution to science* and *public engagement*. Positioning of individual projects is achieved by taking the mean of the scores awarded in each sub-category, while the axes themselves are positioned so that they cross at the mean level of performance observed within the sample. The size of the marker corresponding to each project is representative of the total number of classifications received such that the relationship between success and the 'scale' of the project can be observed.

Figure 3: Project Positioning Matrix



The success matrix appears to show a positive trend relating the positioning of the projects and the size of the marker, indicating that projects receiving more classifications tend to be more successful. Comparatively few projects demonstrate markedly higher levels of public engagement relative to scientific impact (and vice versa) which indicates that the elements of our success matrix are strongly linked (correlation coefficient = +0.54). Galaxy Zoo 4 is the only project that scores relatively well against *public engagement*, but less well against *contribution to science*. Conversely, there is a more significant cluster of projects that observe an opposite relationship; namely Cell Slider, the Milky Way Project and Ancient Lives. This indicates that projects are more likely to make a strong contribution to science despite low public engagement than the reverse.

Table 2 presents coefficients of correlation between the both matrix elements (*contribution to science* and *public engagement* respectively) and the constituent performance indicator measurements used to position the projects. It is clear that the strength of correlation varies between the core elements and the individual components; in some cases quite considerably. Some of the stronger correlations suggest that a more limited subset of these component indicators might do approximately as good a job of explaining the final position on the matrix as the aggregated core element, particularly with respect to *public engagement*. The measurements with lower correlations, e.g. *effective training* or *collaboration*, are those where greater levels of variation are observed between projects, including many zero scores. Those with higher correlations reflect the component indicators where performance was relatively more uniform across the selection of projects, which is to be expected.

Table 2: Correlation between matrix components and individual performance indicators

Matrix Element	Performance Indicator	Measurement	Correlation with Matrix Element
Contribution to Science	Data Value	Publication Rate	0.656
		Completeness of Analysis	0.707
		Academic Impact	0.647
	Project Design and Resource Allocation	Resource Savings	0.572
		Distribution of Effort	0.260
		Effective Training	0.077
Public Engagement	Dissemination and Feedback	Collaboration	0.359
		Communication	0.897
		Interaction	0.869
	Participation and Opportunities for Learning	Project Appeal	0.913
		Sustained Engagement	0.662
		Public Contribution	0.799

We also investigate the robustness of our relative positioning of projects by systematically removing each performance indicator measurement one-by-one from the calculation of the matrix element scores. The results of this analysis are summarised in Table 3, which presents raw numerical scorings for each of the matrix elements (with rank ordering in parentheses) after excluding each individual performance indicator measure. When we undertake this sensitivity analysis, we generally observe greater levels of stability in the ranking of projects occupying the top and bottom ranks, indicating that our measures seem to do a better job of consistently identifying better and less-well performing projects than those in the middle of the distribution. We also find that there is no single measurement that can be removed while preserving the rank order of the projects against their respective matrix elements. Significant variation in the rank ordering of projects is observed even when removing the measurement with the lowest correlation coefficient (*effective training*) from the calculation of the *contribution to science* score; only five out of seventeen projects remain consistently ranked following the recalculation. This leads us to conclude that each individual element of the performance matrix represents an important determinant of the overall positioning of projects, while further demonstrating the need to incorporate a broad mix of indicators that capture different aspects of project performance.

Table 3: Numerical performance indicator measures (rankings) and sensitivity analysis

Project	Contribution to Science	Excluding Publication Rate	Excluding Completeness of Analysis	Excluding Academic Impact	Excluding Resource Savings	Excluding Distribution of Effort	Excluding Effective Training
Galaxy Zoo 1	0.684 (1)	0.621 (4)	0.621 (1)	0.654 (2)	0.621 (1)	0.766 (1)	0.821 (1)
Galaxy Zoo 2	0.609 (5)	0.558 (8)	0.531 (5)	0.647 (3)	0.531 (5)	0.656 (3)	0.731 (4)
Galaxy Zoo 3	0.418 (11)	0.485 (9)	0.301 (12)	0.482 (10)	0.309 (12)	0.428 (10)	0.501 (8)
Galaxy Zoo 4	0.431 (10)	0.441 (12)	0.319 (11)	0.517 (9)	0.319 (11)	0.471 (9)	0.517 (7)
Moon Zoo	0.335 (15)	0.401 (14)	0.207 (16)	0.401 (15)	0.245 (14)	0.351 (15)	0.401 (15)
Planet Four	0.477 (8)	0.572 (7)	0.400 (8)	0.572 (7)	0.382 (9)	0.494 (8)	0.441 (11)
Planet Hunters	0.650 (2)	0.601 (6)	0.581 (3)	0.607 (5)	0.582 (2)	0.752 (2)	0.780 (2)
Solar Stormwatch	0.412 (12)	0.464 (10)	0.328 (10)	0.444 (12)	0.326 (10)	0.416 (11)	0.495 (9)
The Milky Way Project	0.613 (4)	0.669 (2)	0.536 (4)	0.536 (8)	0.548 (4)	0.654 (4)	0.736 (3)
Bat Detective	0.185 (16)	0.223 (16)	0.211 (15)	0.223 (16)	0.193 (16)	0.189 (16)	0.076 (17)
Seafloor Explorer	0.362 (13)	0.434 (13)	0.244 (14)	0.434 (13)	0.248 (13)	0.376 (13)	0.434 (13)
Snapshot Serengeti	0.504 (7)	0.604 (5)	0.404 (7)	0.604 (6)	0.410 (8)	0.551 (5)	0.447 (10)
Whale FM	0.445 (9)	0.454 (11)	0.335 (9)	0.462 (11)	0.523 (6)	0.362 (14)	0.534 (6)
Ancient Lives	0.619 (3)	0.711 (1)	0.600 (2)	0.742 (1)	0.574 (3)	0.542 (6)	0.542 (5)
Cyclone Center	0.147 (17)	0.176 (17)	0.154 (17)	0.176 (17)	0.093 (17)	0.105 (17)	0.176 (16)
Cell Slider	0.524 (6)	0.628 (3)	0.452 (6)	0.628 (4)	0.462 (7)	0.531 (7)	0.440 (12)
Old Weather	0.357 (14)	0.367 (15)	0.273 (13)	0.429 (14)	0.236 (15)	0.411 (12)	0.429 (14)

Project	Public Engagement	Excluding Collaboration	Excluding Communication	Excluding Interaction	Excluding Project Appeal	Excluding Sustained Engagement	Excluding Public Contribution
Galaxy Zoo 1	0.557 (3)	0.622 (3)	0.590 (3)	0.493 (3)	0.493 (3)	0.659 (2)	0.559 (2)
Galaxy Zoo 2	0.649 (1)	0.779 (1)	0.714 (2)	0.579 (2)	0.586 (1)	0.692 (1)	0.680 (1)
Galaxy Zoo 3	0.152 (7)	0.182 (7)	0.167 (7)	0.110 (9)	0.135 (7)	0.178 (7)	0.165 (7)
Galaxy Zoo 4	0.416 (4)	0.387 (4)	0.401 (5)	0.440 (4)	0.299 (4)	0.494 (4)	0.469 (4)
Moon Zoo	0.015 (17)	0.018 (16)	0.017 (17)	0.016 (17)	0.007 (16)	0.018 (17)	0.015 (17)
Planet Four	0.250 (6)	0.299 (5)	0.275 (6)	0.239 (6)	0.193 (6)	0.293 (6)	0.196 (6)
Planet Hunters	0.382 (5)	0.258 (6)	0.320 (4)	0.434 (5)	0.282 (5)	0.456 (5)	0.454 (5)
Solar Stormwatch	0.027 (15)	0.010 (17)	0.019 (15)	0.029 (13)	0.028 (12)	0.032 (15)	0.032 (13)
The Milky Way Project	0.043 (12)	0.052 (12)	0.048 (11)	0.038 (11)	0.029 (11)	0.048 (12)	0.045 (12)
Bat Detective	0.041 (13)	0.049 (13)	0.045 (12)	0.027 (14)	0.043 (10)	0.047 (13)	0.031 (14)
Seafloor Explorer	0.118 (8)	0.141 (8)	0.129 (9)	0.131 (7)	0.059 (8)	0.135 (8)	0.117 (8)
Snapshot Serengeti	0.614 (2)	0.737 (2)	0.676 (1)	0.635 (1)	0.575 (2)	0.537 (3)	0.537 (3)
Whale FM	0.019 (16)	0.023 (15)	0.021 (16)	0.023 (16)	0.007 (17)	0.022 (16)	0.018 (16)
Ancient Lives	0.071 (10)	0.086 (10)	0.078 (10)	0.078 (10)	0.013 (15)	0.084 (10)	0.083 (10)
Cyclone Center	0.030 (14)	0.036 (14)	0.033 (13)	0.026 (15)	0.020 (14)	0.034 (14)	0.028 (15)
Cell Slider	0.102 (9)	0.123 (9)	0.113 (8)	0.123 (8)	0.023 (13)	0.120 (9)	0.102 (9)
Old Weather	0.049 (11)	0.059 (11)	0.054 (14)	0.031 (12)	0.056 (9)	0.058 (11)	0.057 (11)

Part of the observed discrepancy in performance between projects may be related to the nature of the subjects that volunteers are asked to classify in each project, where both Whale FM and Bat Detective involve use of audio clips. This may be indicative that online citizen science projects involving visual tasks are more likely to be successful compared with projects based on other sensory inputs. It should also be noted that the upper-right quadrant of the matrix is predominantly made up of astronomy projects, especially the various incarnations of Galaxy Zoo and Planet Hunters, whereas non-astronomy projects such as Cyclone Center and Bat Detective score comparably less favourably. The only non-astronomy project to feature in the upper-right quadrant is Snapshot Serengeti, which earns its position on the basis of very strong levels of *public engagement*. While the data do not suggest that astronomy projects are the only type that can enjoy success, they do show that astronomy projects tend to score consistently more highly against multiple dimensions or measures of success as opposed to simply one or two. No project exemplifies this interrelationship quite like the original Galaxy Zoo, which combines an active community of volunteers and an engaged science team with a significant number high quality co-authored papers that have had a considerable academic impact.

The apparent dominance of Galaxy Zoo 1 raises a number of important questions about the extent to which astronomy projects may or may not be inherently more suited to the online environment. Was the success of the original Galaxy Zoo simply a result of the novelty of the project at the time it was released in 2007? Are we simply observing a lag in scientific output due to the need for new teams to process new data? Or has a finite pool of volunteer labour become increasingly stretched as the number of new projects increases over time? An analysis of the rate of growth of both new Zooniverse projects and numbers of volunteers shows that the two have increased at much the same rate from 2011-2014, with a geometric mean annual growth in users of 32% compared with 38% annual growth in new projects. This suggests that the Zooniverse has not yet reached a meaningful limit in the pool from which it draws its volunteer resources.

The high proportion of astronomy projects in the upper-right hand quadrant of the matrix may also be a consequence of the Zooniverse platform being founded on the original Galaxy Zoo project and later expanding into a broader suite of ecology and humanities subjects. As a consequence of its history, a significant proportion of the core Zooniverse management team have a formal background in astronomy[5] and many are themselves a part of the project science teams for Galaxy Zoo, Planet Hunters and Milky Way projects. *This may lead to a situation where the central Zooniverse management team have a better understanding of the requirements associated with astronomy research and a greater ease with which projects can be designed to meet these needs. There have also been opportunities to transfer*

[5] A comparison of the central Zooniverse team list appearing on https://www.zooniverse.org/team and the Galaxy Zoo team list appearing on http://www.galaxyzoo.org/#/team shows the degree of overlap between the two groups.

knowledge between science teams of astronomy projects, e.g. the Planet Hunters team benefited from the experiences of Galaxy Zoo in a way that projects like Snapshot Serengeti did not. In an attempt to address this issue, the Zooniverse has already begun a process of recruiting people with more diverse backgrounds.

Finally, despite our efforts to counter the bias towards older projects as much as possible by squaring project age or active duration in our calculations, it still undoubtedly remains the case that older projects have simply enjoyed more time for collaboration and publication compared with their more recently established counterparts. In particular, later iterations of Galaxy Zoo avoided the learning curve faced by new science teams and were formed in the midst of an already successful community. The composition of the Zooniverse has changed dramatically over the last four years; of the 18 considered projects created during or before 2010, 7 of 8 were astronomy projects, while only 3 out of the 10 projects created after 2010 were related to astronomy. Developments in the Zooniverse over the coming years will therefore allow for a more decisive assessment of whether astronomy projects really are inherently more successful than others, or whether there is simply a delay in other projects 'catching up' with the early movers from astronomy.

For other organisers of citizen science projects, the implications of these findings would be to first recognise the importance and the strength of relationship between scientific impact and public engagement. We show it is relatively unlikely for a citizen science project to meet with success against one of these measures and not the other, so an effective management strategy should target the achievement of both goals instead of one in isolation. Second, organisations overseeing a number of online projects relating to different areas of science (such as Crowdcrafting) should carefully assess their core scientific competencies and the effect this may have on the design and success of new projects, as this may affect the extent to which diversification is required among the core management team. Finally, as many of the project outcomes and performance measures considered in this paper relate to activities overseen by project science teams, there is a need to pay careful attention to the knowledge and training of the scientists running projects. A well-made interface alone does not appear to be a sufficient condition to achieve a successful outcome.

The literature on citizen science is still nascent. Although this framework makes an important contribution with relevance across disciplines, there are still a number of limitations that may also guide future research undertaken in this area. First, we have been limited to some extent by the practical availability of performance data that may have been useful to supplement or enhance the measures of project success. These include the scale of the project measured in terms of the amount of resources at its disposal, the level of training provided to the project team and information on the socio-economic diversity of project participants to name a few. Second, our study is limited to an extent by the inherent subjectivity of defining and interpreting success. While we have used the literature on citizen science to define this term in an objective sense, success is in reality a highly nuanced concept. The appropriate definition of success might vary greatly depending on the parties involved in a project and the unique goals they have set. For instance, although some of our criteria measure publication and citation counts, it is entirely plausible that a project might be considered successful by those involved even if it doesn't result in any academic publications. Future studies may be able to build on this by work by making use both of additional data and by combining the 'bottom up' aggregated approach outlined in this study with a mixed-methods approach incorporating qualitative data from surveys or interviews. A combination of these extensions would provide an additional depth of insight to complement the work presented here.

5. Conclusion

This study has presented a unique framework for measuring and assessing success in online citizen science projects. We argue that successful projects are those that achieve high scientific impact as well as high levels of public engagement and that four key elements can be used to score projects against these criteria; namely *data value*; *project design and resource allocation; dissemination and feedback* and *participation and opportunities for learning*. Performance data are collected from a sample of 17 online citizen science projects forming part of the Zooniverse platform and are scored in order to position these projects on a matrix of relative success. The results demonstrate that scientific impact and public engagement are positively correlated and a high proportion of the most successful projects are related to the field of astronomy. These results have significant implications for the management and organisation of citizen science projects, namely that the objectives of scientific impact and public engagement need to be considered jointly rather than separately, since projects are seemingly less likely to be successful if performing well against only one of these measures. A broader issue arising from the case study of the Zooniverse is the effect of core competencies and expertise of the central management team on the likelihood of success for projects in particular subject areas. In order to address this particular issue, the Zooniverse has already begun taking measures to diversify the background and expertise of their staff.

Acknowledgements

We would like to acknowledge the Engineering and Physical Sciences Research Council (EPSRC) and New Economic Models in the Digital Economy (NEMODE) Network+ for funding this research as part of the Volunteer and Crowdsourcing Economics (VOLCROWE) project. We would also like to thank the Zooniverse for their co-operation.

References

[1] Silvertown, J. (2009). A new dawn for citizen science. Trends in Ecology & Evolution, 24(9), 467-471.

[2] Greenhill A. Holmes, K. Lintott, C. Simmons, B. Masters, K. Cox, J. & Graham, G. (2014). Playing with Science: Gamised Aspects of Gamification found on the Online Citizen Science Project - Zooniverse. In: Dickinson P. (ed) GAME-ON 2014. University of Lincoln, UK, pp. 15 - 24.

[3] European Commission (2013). Science for environmental policy in-depth report: Environmental citizen science: Science Communication Unit, University of the West of England Bristol. Retrieved from: http://ec.europa.eu/environment/integration/research/newsalert/pdf/IR9.pdf

[4] Council of Canadian Academies (2014). Science Culture: Where Canada Stands: The Expert Panel on the State of Canada's Science Culture, Council of Canadian Academies. Retrieved from: http://www.scienceadvice.ca/uploads/eng/assessments%20and%20publications%20and%20news%20releases/science-culture/scienceculture_fullreporten.pdf

[5] National Research Council of the National Academies (2010). New Worlds, New Horizons in Astronomy and Astrophysics. National Academies Press. Retrieved from: http://science.nasa.gov/astrophysics/special-events/astro2010-astronomy-and-astrophysics-decadal-survey/

[6] Fortson, L., Masters, K., Nichol, R., Borne, K.D., Edmondson, E.M., Lintott, C., Raddick, J., Schawinski, K. & Wallin, J. (2012). Galxy Zoo: Morphological classification and citizen science in: Advances in Machine Learning and Data Mining for Astronomy, CRC Press, Taylor & Francis Group, Eds.: Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, Ashok N. Srivastava, p. 213-236.

[7] Lintott, C.J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M.J., Nichol, R.C., Szalay, A., Andreescu, D., Murray, P. & Vandenberg, J. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. Monthly Notices of the Royal Astronomical Society, 389(3), 1179-1189.

[8] Wiggins, A. & Crowston, K. (2010). Developing a conceptual model of virtual organisations for citizen science. International Journal of Organisational Design and Engineering, 1(1), 148-162.

[9] Wiggins, A. & Crowston, K. (2011, January). From conservation to crowdsourcing: A typology of citizen science. In System Sciences (HICSS), 2011 44th Hawaii International Conference on (pp. 1-10). IEEE.

[10] Wiggins, A. & Crowston, K. (2012, January). Goals and tasks: Two typologies of citizen science projects. In System Science (HICSS), 2012 45th Hawaii International Conference on (pp. 3426-3435). IEEE.

[11] Cohn, J. P. (2008). Citizen science: Can volunteers do real research? BioScience, 58(3), 192-197.

[12] Cashman, S. B., Adeky, S., Allen III, A. J., Corburn, J., Israel, B. A., Montaño, J. & Eng, E. (2008). The power and the promise: working with communities to analyze data, interpret findings, and get to outcomes. American Journal of Public Health, 98(8), 1407-1417.

[13] Gardiner, M. M., Allee, L. L., Brown, P. M., Losey, J. E., Roy, H. E. & Smyth, R. R. (2012). Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. Frontiers in Ecology and the Environment, 10(9), 471-476.

[14] Sheppard, S. A. & Terveen, L. (2011, October). Quality is a verb: The operationalization of data quality in a citizen science community. In: Proceedings of the 7th International Symposium on Wikis and Open Collaboration (pp. 29-38). ACM.

[15] Lortie., C.J., Aarssen, L.W, Budden, A.E. & Leimu, R. (2013). Do citations and impact factors relate to the real numbers in publications? A case study of citation rates, impact and effect sizes in ecology and evolutionary biology. Scientometrics, 94, 675-682.

[16] Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V. & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience, 59(11), 977-984.

[17] Raddick, M. J., Bracey, G., Carney, K., Gyuk, G., Borne, K., Wallin, J. & Jacoby, S. (2009). Citizen science: status and research directions for the coming decade. AGB Stars and Related Phenomenastro 2010: The Astronomy and Astrophysics Decadal Survey, 46P.

[18] Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C. & Jacobs, D. (2012, February). Dynamic changes in motivation in collaborative citizen-science projects. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (pp. 217-226). ACM.

[19] Riesch, H. & Potter, C. (2013). Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. Public Understanding of Science, 0963662513497324.

[20] Franzoni, C. & Sauermann, H. (2014). Crowd science: The organization of scientific research in open collaborative projects. Research Policy, 43(1), 1-20.

[21] Dai, P. & Weld, D. S. (2010, April). Decision-theoretic control of crowd-sourced workflows. In Twenty-Fourth AAAI Conference on Artificial Intelligence.

[22] Bauer, M. W. & Jensen, P. (2011). The mobilization of scientists for public engagement. Public Understanding of Science, 20(1), 3.

[23] Newman, G., Zimmerman, D., Crall, A., Laituri, M., Graham, J. & Stapel, L. (2010). User-friendly web mapping: lessons from a citizen science website. International Journal of Geographical Information Science, 24(12), 1851-1869.

[24] Brossard, D., Lewenstein, B. & Bonney, R. (2005). Scientific knowledge and attitude change: The impact of a citizen science project. International Journal of Science Education, 27(9), 1099-1121.

[25] Cronje, R., Rohlinger, S., Crall, A. & Newman, G. (2011). Does participation in citizen science improve scientific literacy? A study to compare assessment methods. Applied Environmental Education & Communication, 10(3), 135-145.

[26] Phillips, T. B., Ferguson, M., Minarchek, M., Porticella, N. & Bonney, R. (2014). User's Guide for Evaluating Learning Outcomes in Citizen Science. Ithaca, NY: Cornell Lab of Ornithology.

[27] Trumbull, D. J., Bonney, R., Bascom, D. & Cabral, A. (2000). Thinking scientifically during participation in a citizen-science project. Science Education, 84(2), 265-275.

[28] Cummings, J. N. & Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. Research Policy, 36(10), 1620-1634.

[29] Hall, K. L., Stokols, D., Stipelman, B. A., Vogel, A. L., Feng, A., Masimore, B. & Berrigan, D. (2012). Assessing the value of team science: a study comparing center-and investigator-initiated grants. American Journal of Preventive Medicine, 42(2), 157-163.

[30] Becher, T., & Trowler, P. (2001). Academic tribes and territories: Intellectual enquiry and the culture of disciplines. McGraw-Hill International.

[31] Shin, J. C., & Cummings, W. K. (2010). Multilevel analysis of academic publishing across disciplines: Research preference, collaboration, and time on research. Scientometrics, 85(2), 581-594.

[32] Holmberg, K. & Thelwall, M. (2013). Disciplinary differences in Twitter scholarly communication. Scientometrics, 1-16.