CITIZEN SCIENCE:
THEORY AND PRACTICE

Serret, H, et al. 2019. Data Quality and Participant Engagement in Citizen Science: Comparing Two Approaches for Monitoring Pollinators in France and South Korea. *Citizen Science: Theory and Practice*, 4(1): 22, pp. 1–12. DOI: https://doi.org/10.5334/cstp.200

RESEARCH PAPER

# Data Quality and Participant Engagement in Citizen Science: Comparing Two Approaches for Monitoring Pollinators in France and South Korea

Hortense Serret*, Nicolas Deguines†, Yikweon Jang*, Grégoire Lois‡ and Romain Julliard‡

Citizen science has become a mainstream approach for collecting data on biodiversity. However, not all biodiversity monitoring programs achieve the goal of collecting datasets that can be used in robust scientific inquiries. Data quality and the capacity to engage participants in the long-term are the most challenging issues. We compared two methodologies of citizen science programs dedicated to pollinators monitoring in France (Spipoll) and South Korea (K-Spipoll). These programs aimed to launch long-term monitoring at a community-level to better understand environmental effects on the composition and stability of pollinator communities. We assessed, through different metrics, how the two approaches influenced (1) data quality (assessed by "Accuracy in data collection," "Consistency in protocol relative to volume of sessions contributed by an individual," "Spatial representation of data," and "Sample size"), and (2) participant engagement (assessed by "the number of connected days," "the number of active days," "the proportion of participant contributing a single session," "the average number of sessions per participant," and "the distribution of numbers of contributions per participant in each program."). On one hand, participants in the Spipoll program abided by the standard protocol more often and provided identification for the photographed insects, leading to efficient ecological analyses. On the other hand, the K-Spipoll program provided more sessions per participant and a lower rate of single participation, with a full session demanding less effort in terms of data input, providing critical data where baseline data have otherwise been unavailable. These differences have emerged through methodology choices: For the Spipoll, the dedicated website favored the emergence of a social network that facilitated identification and increased data quality; for the K-Spipoll, the development of a cell phone application facilitated participation, and regular on-field education sessions motivated participants. We conclude by providing suggestions for the implementations of future citizen science programs to improve both data quality and participant engagement.

**Keywords:** Citizen scientists; Pollinator survey; Data quality; Participants engagement; Digital application; Participation

## Introduction

Citizen science, defined as participation of the general public in scientific research, has become a mainstream approach for collecting data on biodiversity (Chandler et al. 2017). This approach can substantially help scientists to address large scale or global biodiversity issues (Chandler et al. 2016; Thornhill et al. 2016), through (1) monitoring the state of biodiversity by bringing a large amount of data needed in macro-ecology analysis (Devictor, Whittaker, and Beltrame 2010; Deguines et al. 2012; Muratet and Fontaine 2015; Olivier et al. 2016) and (2) the creation of indicators providing relevant information about the state of biodiversity and public concern and action (Couvet et al. 2008; Jiguet et al. 2012, French Biodiversity Observatory 2018). The development of citizen science is an opportunity for the general public to familiarize with scientific thinking (Trumbull et al. 2000) and to improve their knowledge on specific subjects (Bonney et al. 2009; Deguines et al. 2018), which has been highlighted as one of the main motivations of participation (Bruyere and Rappe 2007; Domroese and Johnson 2017; Curtis 2018). More specifically, citizen science is also a way to raise societal awareness about the stakes of biodiversity conservation (Lewandowski and Oberhauser 2017). The participation in nature-based citizen science projects can also be a way to reconnect people with nature (Guiney and Oberhauser 2009).

Recently, many citizen science programs monitoring pollinators have been launched (e.g., Deguines et al. 2012; Domroese and Johnson 2017; Suzuki-Ohno et al. 2017)]. This trend is tied to the growing awareness of pollinator

* Ewha Womans University, KR

† Université Paris-Sud/CNRS/AgroParisTech, FR

‡ Muséum national d'Histoire naturelle, Centre National de la Recherche Scientifique, Sorbonne Université, FR

Corresponding author: Hortense Serret (hortense.serret@gmail.com)

declines (Potts et al. 2010), which are a major threat to the functioning of terrestrial ecosystems. Indeed, 87.5% of angiosperms depend on animal pollination (Ollerton, Winfree, and Tarrant 2011). To help policy-makers meet the challenge of pollinator conservation, indicators of community composition changes and population trends are needed, triggering the set-up of long-term monitoring programs. Participation of the general public in monitoring pollinators plays a critical role because many data over large areas and multiple years (even decades) is required (Dickinson, Zuckerberg, and Bonter 2010).

Some programs successfully addressed specific questions regarding the biology of a target species such as the Monarch butterfly (Ries and Oberhauser 2015), the distribution of bumblebees in Japan (Suzuki-Ohno et al. 2017), community-level responses to land-use changes (Deguines et al. 2012; Deguines et al. 2016; Desaegher et al. 2018; Levé, Baudry, and Bessa-Gomes 2019), or global-scale mapping of pollination services (LeBuhn, et al. 2016). Silvertown et al. (2013) reported that, thanks to a citizen science project conducted in the UK, two species of insects never recorded in the country were discovered "*including the first record of the Euonymus leaf notcher moth, discovered by a 6-year-old girl*."

However, the credibility of citizen science is sometimes debated. Data quality is one of the most challenging issue about citizen science (Cohn 2008; Bird et al. 2014), and some authors point out that many projects fail to provide data of sufficient quality for publishing in peer-rewiewed scientific journals (Theobald et al. 2015).

There are multiple aspects to data quality such as precision and accuracy in data collection, consistency in protocol between individuals and over time, adequate spatial and temporal representation, and sufficient sample size for statistical inferences (Lewandowski et al. 2015), all of which must be considered if researchers are to answer the research questions they seek to address. Many systematic methods can be used to ensure high levels of data quality, such as designing standard protocols adapted to participant's skills and research questions, participant training, cross-checking, validation of observations and identifications by experts or other participants, systematic screening for aberrant contributions, technological help (automatic sorting or identification), and others (Wiggins et al. 2011; Freitag, Meyer, and Whiteman 2016).

Data quality is also linked to participant's engagement through two mechanisms. First, engaged participants may contribute more observations and be more involved in the long-term: Collectively, overall quality of the dataset gathered thus increases and allows investigation of more complex (including temporal) questions. Second, as a citizen scientist continues to participate in a monitoring program, skills and knowledge may improve and resulting quality of individual data may be heightened (Deguines et al. 2018). Successfully engaging people in citizen science programs thus represent a critical challenge.

In this study, we compared methodologies of two citizen science programs launched in France and South Korea. The data collected are used to address conservation and research questions related to pollinator distribution, richness, community composition, and community stability, and to understand how these characteristics are influenced by landscape composition, connectivity, and management from a long-term perspective. These objectives require standardized protocols allowing collection of many data about the presence of all the species observed on different types of landscapes over multiple years. Although both programs relied on similar protocols to monitor pollinators using photographs, there were a few differences in terms of methods and management. We assessed how the two approaches influenced (1) data quality and (2) participant's engagement. Multiple metrics were used to assess data quality and level of participant engagement of both programs based on criteria both published (Lewandowski and Specht 2015; Ponciano and Brasileiro 2014) and developed by us.

We discuss how differences in metrics between the two programs may arise from their functioning and methodology. From the critical assessment of the advantages and limits of these two case studies, we aim to emphasize key methodological choices in the development of citizen science programs using digital tools.

## Methods

### *The Photographic Survey of Flower Visitors, Spipoll*

The Photographic Survey of Flower Visitors, hereafter called Spipoll, was launched in France in 2010 (Deguines et al. 2012). The program was established by the National Museum of Natural History of Paris (France) in partnership with the Office for insects and their environment (Opie, an entomological society). A website especially developed for the program provides information about the critical roles of pollinators for ecosystem functioning and the importance of long-term monitoring in scientific investigations (http://www.spipoll.org).

The standard protocol asks participants to choose a flowering plant species and to take pictures of all invertebrates landing on its flowers during a 20-minute period. The observations can be done on an area of 10 m², as long as all of the pictures were taken on the flowers belonging to the same plant species. Observations can be done wherever participants can find a flowering plant (from dense urban centers to natural areas). Participants were also asked to take pictures of the plant and its environment and to provide date and time information, Global Positioning System coordinates, habitat characteristics, and climatic conditions (wind, temperature, cloud cover). Written tutorials explaining the protocol in detail were available on the website.

After their observations, participants had to sort their pictures and keep a single picture per species, choosing what they felt was the most useful one for insect identification. The website would not allow a participant to upload a session if s/he had not tried to identify at least 50% of the photographed insects. Participants then identified the plant and insects using online computer-aided identification tools especially developed for the Spipoll (Deguines et al. 2012). These tools allowed observers to identify pollinators and plants using descriptors related to morphological traits (e.g., length of antennas, eyes shape, color pattern; number and color of petals), choosing among 556 insects or insect groups and 333 plant

morphospecies. All descriptors are explained through text and illustrations, using pictures featuring different examples. Entomologists and botanists review insect and plant pictures and correct the identification when necessary. Each set of pictures of insects and associated plants from 20 minutes of observation at a given date and place is hereafter referred as a "session."

A strong community management was set up through the dedicated website. Some entomologists from the Opie provided comments on the posted observations. Any participants could also comment on observations from others and notify them of potential incorrect identification or misapplied protocol. Eventually, a social network emerged from the program, involving both community managers from the Opie and observers themselves. In addition, observers received a monthly newsletter that provided information on the progress in overall participation, highlighted a "plant of the month" in bloom, featured a rich session from one participant, and shared interesting facts about pollination.

### *The Korean Photographic Survey of Pollinators, K-Spipoll*

In 2017, the Korean Photographic Survey of Pollinators (hereafter abbreviated "K-Spipoll") was launched in partnership with a publishing company (Donga Science) which proposed to its subscribers to participate in ecological surveys on different taxa (plants, cicadas, birds, treefrogs). This program, called "The Earth Lovers Explorers," is the first initiative of citizen science in South Korea and was established in partnership with researchers from Ewha Womans University to address research questions in ecology and conservation biology. The protocol of K-Spipoll was the same as that of Spipoll, except that participants were asked to conduct the survey on a 15-minute period. The following metadata were collected with pictures: Date, time of day, GPS coordinates, and environmental conditions.

The data were collected through a dedicated cellphone application developed by Donga Science. This digital application was open only to its subscribers. Consequently, the targeted observers were the readers of the magazine (i.e., children and their parents).

Identification of the insects was not asked of observers of the K-Spipoll. The pictures were uploaded on the website and identified by professionals of Ewha Womans University, scientific partners of the company, and amateur entomologists. Identifications were validated by an entomology expert, Dr. Lee Heung-Sik from Plant Quarantine Technology Center.

The publishing company and Ewha Womans University organized training events for K-Spipoll participants. Six training sessions (in which participation was optional) were organized between April and May 2017 in several parks of Seoul; 60% of the first-year participants came to one of these sessions, and 83% of those who came uploaded data. The researchers of Ewha Womans University presented the importance of pollinators in ecosystems and the need to protect them, and highlighted the advantages of citizen science programs for understanding pollinator communities. The researchers showed the participants how to carry out the protocol and trained them in doing it. Experts encouraged and welcomed questions from the participants about the protocol or pollinator ecology in general.

### *Assessing data quality and participant's engagement*

According to a systematic review of the peer-reviewed literature, Lewandowski et al. (2015) identified four aspects of data quality: "Data collection" (precision and accuracy in data collection); "Standardized sampling" (consistency in protocol between individuals and over time); "Spatial and temporal representation" (adequate spatial and temporal representation); and "Sample size" (sufficient sample size for statistical inferences). We took inspiration from these four aspects and adapted them to the Spipoll and the K-Spipoll's specificities to assess data quality in relation to the programs' research questions.

Our first metric, "Accuracy in data collection," was measured through the proportion of sessions respecting the protocol. Indeed, the accuracy of the information collected, i.e., richness of pollinators and community composition observed in a given time, is necessary to address the research questions. We considered the protocol to have been violated when a session was not georeferenced, included pictures taken on different plant species or on leaves, because pictures across multiple plant species cannot be assumed to represent community or abundance of insects using a single plant species over a fixed temporal period. Likewise, photos of leaves cannot be assumed to represent the community or abundance of pollinators associated with flowers. Furthermore, we also considered sessions containing only a single insect picture as violating the protocol. The probability to observe only one species in 15 or 20 minutes is low (see Supplementary File); instead, this likely occurred if a participant did not observe during the 15 minute or 20 minute period or if there were a misunderstanding in the data that required uploading (e.g., upload one photo of an insect observed during the session). This violation would lead to an underestimation of the pollinator's richness and would introduce a bias when analyzing variations in pollinator communities, because the observation effort would not be standardized and thus comparable between the sessions. For the rest of our analysis, the sessions considered as "strict violation of protocol" were removed from the dataset because they were not exploitable with regard to plant-pollinator ecology nor our analysis.

Our second metric, "Consistency in protocol relative to volume of sessions contributed by an individual," was measured by comparing the proportion of single-species sessions according to participation; i.e., between participants having done one session, from two to 10 sessions, and more than 10 sessions. This metric assessed the point at which participation leads to a better understanding of the protocol, thanks to, for instance, the social networks created around the programs.

Our third metric, "Spatial representation of data," was assessed by using two metrics of different spatial scales: Overall dataset level and participant's level. First, we evaluated whether participation within administrative regions of France and South Korea were proportional to population size. Such a spatial distribution appears as a reasonable

objective for citizen science programs, corresponding to collecting data primarily from where people live. Through geographical data processing or focusing on densely populated ecosystems, such dataset were shown to be scientifically valuable, especially to study the relationship between land use and biodiversity (Deguines et al. 2012, 2016; Desaegher et al. 2018; Levé, Baudry, and Bessa-Gomes 2019). Specifically, we computed linear regressions between the number of sessions and the population size in the 22 and 16 administrative regions of France and South Korea respectively, after both variables were *ln(variable + 1)* transformed. We used the *rho* coefficients retrieved from Spearman correlation tests and the coefficient of determination $R^2$ to compare the two regressions.

As a second spatial metric, we assessed the data spatial dispersion for each participant who did at least three sessions. We assumed that the more "exploratory" participants are, the less the data have the risk to be spatially auto-correlated. For each participant, we calculated the median distance of each session to the centroid of his/her sessions. We then performed a Wilcoxon test to evaluate whether participants from South Korea and France differed in their spatial dispersion of participation. For this test, sample size was the number of participants from both programs who did three sessions or more (i.e., n = 165, with 95 and 70 participants for the Spipoll and the K-Spipoll respectively).

The metric "Sample size" was assessed by the total number of sessions done for each program. The size of this metric determines the statistical power for data analyses. Moderately large datasets (e.g., 1000–3000 sessions distributed across a broad geographical areas) allowed investigating macro-ecological dynamics of pollinator communities (e.g., Deguines et al. 2016, Levé, Baudry, and Bessa-Gomes 2019). However, species distribution modelling for a given species of interests could be done with as little as ca. 100 records (e.g., Le Féon et al. 2018 investigating the range expansion of the exotic *Megachile sculpturalis* in France, using records from various sources).

Metrics used for assessing participant's engagements were proposed by Ponciano and Brasileiro (2015) and measure the involvement and interaction of participants with a project over time. For our study, we used two of their metrics. First, we counted the number of days between the first and the last observation (hereafter "connected days"), which represents the amount of time that participants remained linked to the program. Second, we assessed the number of active days (number of days with one participation or more), representing the motivation of participants to participate several times in the year rather than participating several times on a single day. These two metrics are critical to increase the sample size and potentially the spatial representation. Wilcoxon tests were used to test whether these two metrics differed between participants in the Spipoll and the K-Spipoll. We further considered four additional metrics of participant's engagement: The number of participants, the proportion of participants contributing a single session, the average number of sessions per participant, and the distribution of numbers of contributions per participant in each program (allowing determination of the proportion of participants contributing to 50% of the data). These last metrics are also linked to the sample size. More particularly, the number of single sessions is further linked to temporal issues. Indeed, single participations are not ideal for assessing temporal trends.

### *Comparing methodologies*

Design, methodology, and functioning specificities of both programs are presented in **Table 1**. After comparing the above-mentioned metrics of the efficiency of the two programs in gathering the correct data and engaging volunteers, we examine and discuss how programs' specificities could lead to the differences observed.

## Results

### *Data quality*

In general, the Spipoll program had better results than the K-Spipoll program (**Table 2**). In regard to accuracy in data collection, we found that 57% of the Spipoll sessions followed the protocol in contrast to only 26% of K-Spipoll sessions. The strict violations of the protocol (non-georeferenced picture, pictures taken on different plant species or on leaves, no insect on the picture) were the most common source of non-respect of the protocol (29% of the sessions for Spipoll and 39% for the K-Spipoll). The proportion of single-species sessions was greater for the K-Spipoll (35%) than for the Spipoll (14%). This highlights the fact that these participants either misunderstood the protocol (which states to upload one picture per insect) or did not observe during the 15 or 20 minute period.

The percentage of sessions containing only one species decreased over time for the participants of the Spipoll. Participants having uploaded more than 10 sessions had only 15.7% of single-species sessions, compared to 31.6% for participants having uploaded only one session (**Figure 1**). In contrast, the amount of single-species sessions with K-Spipoll did not decrease with the volume of sessions contributed by an individual.

The spatial distribution of sessions was more highly correlated to the spatial distribution of population in the Spipoll program ($R^2 = 0.82$; p-value < 0.0001) than in the K-Spipoll program ($R^2 = 0.49$; p-value = 0.0025) (**Figure 2**).

Conversely, observations by K-Spipoll participants were more spread out in the landscape (median distance = 15.3 km) than observations by Spipoll observers (median distance = 1.9 km; Wilcoxon tests associated p-values p-value < 2.2e–16, Z = –13.16) (**Figure 3**).

### *Participant engagement*

K-Spipoll participants demonstrated greater engagement across metrics than Spipoll participants (**Table 2**). The number of connected days and the number of active days were greater for the participants of the K-Spipoll program with 70.8 (±66.5) connected days in average (n = 87) against 17.2(±24.5) for the Spipoll program (n = 417) (**Figure 4**; Wilcoxon tests associated p-values < 2.2e–16, Z = –8.15 and Z = –8.31 respectively).

With 75% fewer participants, the K-Spipoll reached almost the same number of sessions as the Spipoll, as

**Table 1:** Comparison of the methodologies for the K-Spipoll and the Spipoll.

| | Spipoll | K-Spipoll |
|---|---|---|
| **Generalities** | | |
| Launching year | 2010 | 2017 |
| Observers targeted | Anyone, general public | Local publishing companies subscribers, children, and their parents |
| Duration of the protocol | 20 min | 15 min |
| **Community management** | | |
| Possibility of comments | Yes | Yes |
| Professional animation of the social network, comments on incorrect participations | Yes | No |
| Newsletter | Yes, monthly | None |
| **Field training** | | |
| In field education and training | No | Yes, 6 times in 2017 |
| **Partners involved** | | |
| Research institution | National Museum of Natural History, Paris | Ewha Womans University, Seoul |
| Partners for community management | Entomological society | Publisher private company |
| **Data collection** | | |
| Photographic material | Digital camera | Smartphone |
| Data upload | Dedicated website | Smartphone application |
| **Identification** | | |
| Species identification by the observers required | Yes | No |
| Identification material | Online interactive identification key – Computer-aided identification tool | Identification information in a guide book |

**Table 2:** Comparison of data quality and participant's engagement between the K-Spipoll and the Spipoll projects.

| | Metrics | Spipoll | K-Spipoll |
|---|---|---|---|
| **Data quality** | Number of sessions | 1,853 | 2,163 |
| | Proportion of sessions following the protocol | 57% | 26% |
| | Proportion of strict violation of protocol; proportion of single-species sessions | 29%; 14% | 39%; 35% |
| | Proportion of participants who did only single-species sessions | 22% | 50% |
| | Correlation between the population in administrative regions and the number of sessions (Spearman's *rho* coefficient) | 0.80 | 0.67 |
| | Median distance to centroid per participant (in km) | 1.9 | 15.3 |
| **Engagement** | Number of participants | 529 | 118 |
| | Average number of connected days | 17.2 (±24.5) | 70.8 (±66.5) |
| | Average number of active days | 2.04 (±1.5) | 7 (±6) |
| | Average number of participations per participant | 2.8 (±2.4) | 13.03 (±12.1) |
| | Proportion of single participation | 60.8% | 17.1% |
| | Proportion of main contributors contributing to 50% of the observations | 10% | 13% |

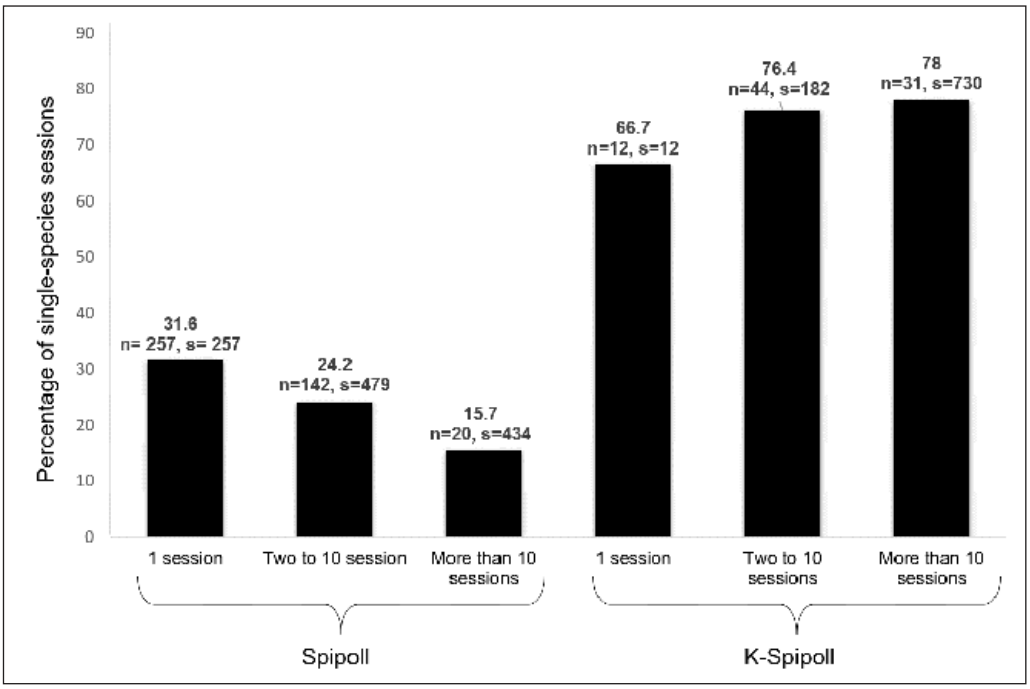


**Figure 1:** Percentage of single-species sessions according to the participant's volume of sessions (where n represents the number of participants and s the sum of sessions).

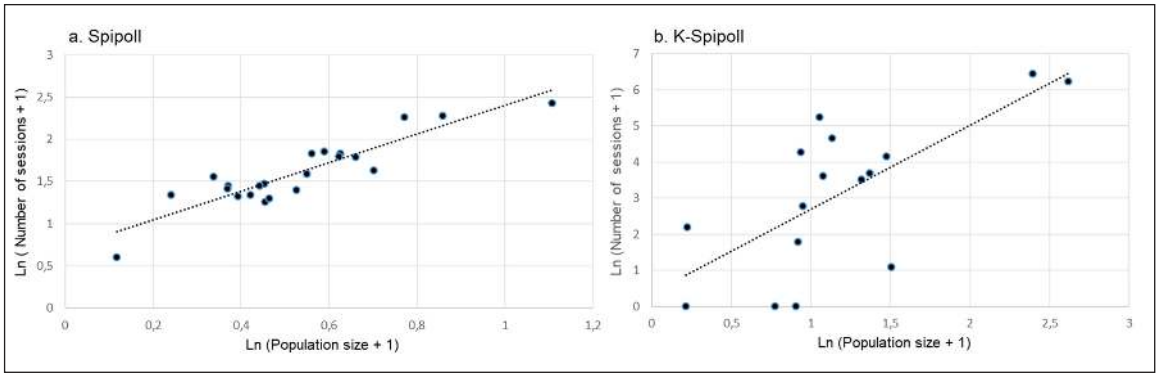


**Figure 2:** Number of sessions according to the population in administrative regions in France **(a)** and South Korea **(b)**.

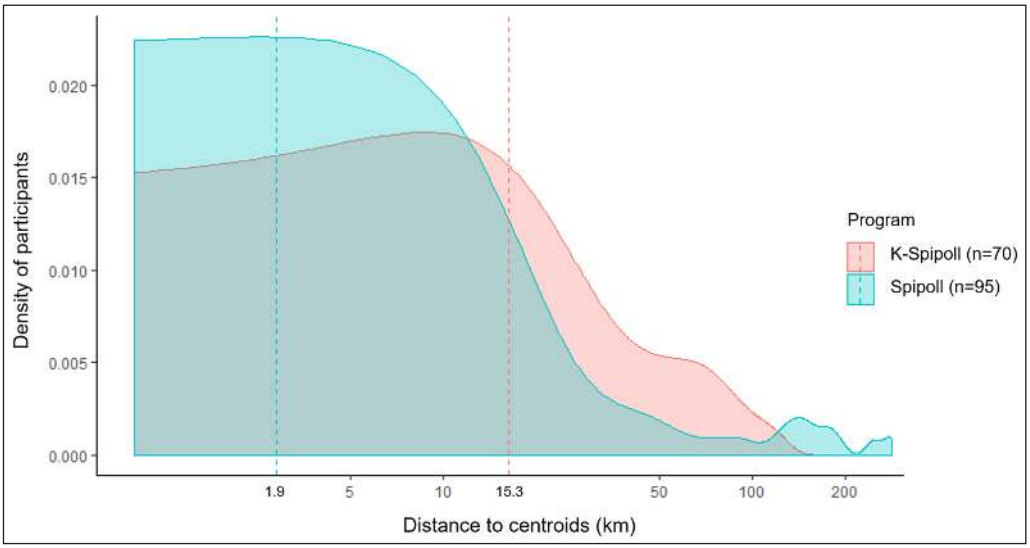


**Figure 3:** Distance to centroids per observers for each program. In each panel, dotted lines represent the median value for Spipoll and K-Spipoll (in blue and red respectively).

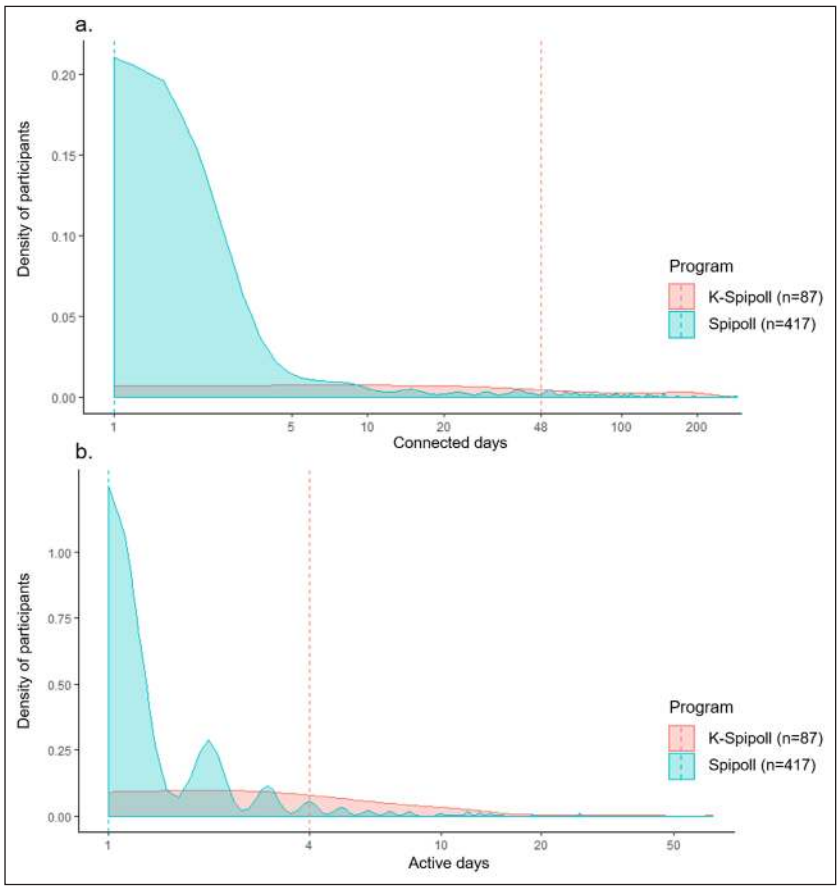


**Figure 4:** Number of connected **(a)** and active **(b)** days per observers for the Spipoll (blue) and the K-Spipoll (red). In each panel, dotted lines represent the median value for Spipoll and K-Spipoll (in blue and red respectively).

K-Spipoll participants each uploaded 13 sessions in average, compared to 2.8 sessions in average for participants of the Spipoll. Additionally, the proportion of single participation is lower in the K-Spipoll (17.1%) than the Spipoll (60.8%).

The number of observations per participant and their contribution to the entire dataset is represented in **Figure 5**, which shows that for K-Spipoll, 13% of participants (i.e., 13 observers) are contributing to 50% of the dataset. These participants did 26 sessions each or more. For the Spipoll, 10% of participants (i.e., 42 observers) collected 50% of the data, each doing 5 sessions or more.

## Discussion

The comparison of Spipoll and K-Spipoll showed that both methodologies had strengths and weaknesses regarding the different metrics we used to assess data quality and participant engagement. The Spipoll program is providing data of high quality regarding specifically the accuracy of data collection and the sample size usable to conduct analyses, with most participants abiding by the standard protocol and providing identification for the photographed insects. As a result, this program successfully published analyses about contrasted affinities of pollinators with different land-use (Deguines et al. 2012), urbanization effects on community composition (Deguines et al. 2016), and more recently works about floral morphology as the main driver of flower-feeding insect occurrences in the Paris region (Desaegher et al. 2018) or the role of domestic gardens as favorable pollinator habitats in impervious landscapes (Levé, Baudry, and Bessa-Gomes 2019). However, the cost of data upload for participant is high and demanding, which negatively influenced participant engagement. For the K-Spipoll, the pictures needed to be sorted and identified by researchers, which limited possibilities for prompt data analyses. Furthermore, the high proportion of single-species sessions for the first year limited the possibility of analyses. For the time being, it is possible that these data may be more challenging to use, or will be useful for a narrower range of questions, because of the quality issues (mainly that observation effort may not have been properly standardized). However, given flower visitor data in South Korea were very scarce before this program, these data constitute critical information on the presence of pollinator species that were not previously available. Some analysis using presence-only data to conduct Species Distribution Modeling and modeling of ecological networks are nevertheless possible where the sample size is large enough.

We showed that consistency in protocol between individuals and over time was progressing for the Spipoll, as the number of single-species sessions decreased after several participations, showing that the participants were understanding the protocol better after several participations, but this was not the case for the K-Spipoll. The K-Spipoll program showed more efficiency for the participants' engagement, a full participation demanding less effort in terms of data input. Participants were "connected" to the

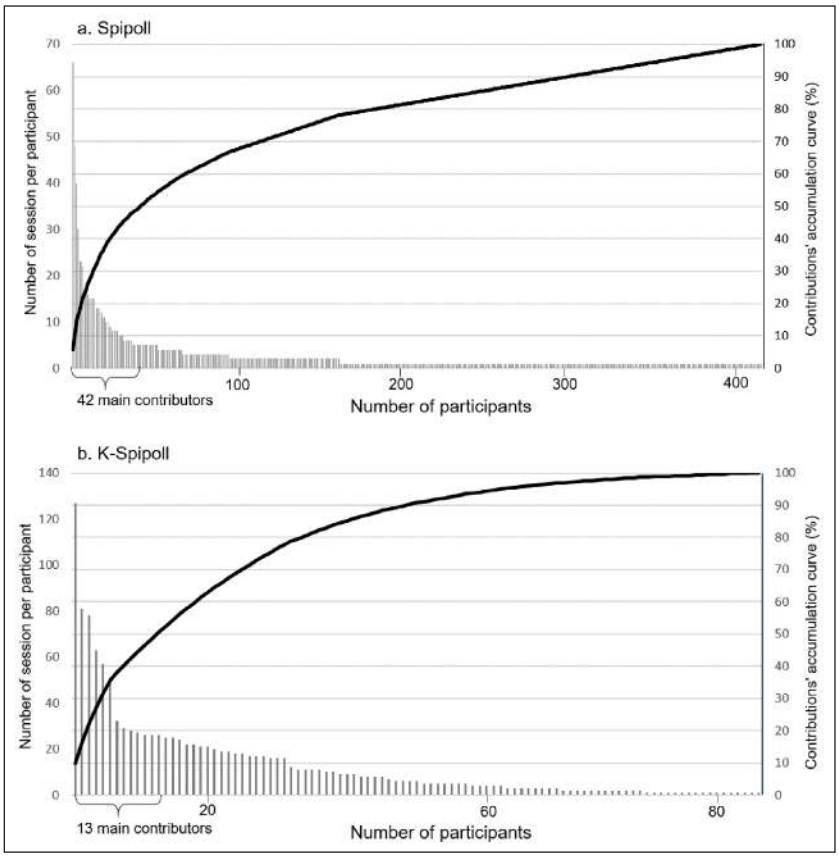


**Figure 5:** Number of sessions per participant (histogram) and contribution to the dataset (accumulation curve) of the Spipoll **(a)** and the K-Spipoll **(b)**.

project for a longer period, were participating more often (number of "active days"), and uploaded more sessions that were also spatially more widespread in K-Spipoll than the Spipoll. A "main contributor" (defined as being among the most active participants contributing to 50% of the dataset) for the K-Spipoll sent at least 26 sessions, whereas such contributors in the Spipoll did five sessions or more. Thus, the proportion of main contributors was slightly highest for the K-Spipoll (13% of participants) than for the Spipoll. The strong commitment of the participants of the K-Spipoll is encouraging in terms of long-term participation and to address the temporal monitoring aims of the program. We discuss below how these differences could have emerged, and provide suggestions for the implementation of future citizen science programs.

### *The social network vs. on-field activities*

We suggest that the community management of the social network dedicated to the Spipoll program drove participant respect of the protocol. Indeed, a community manager provided online personalized and constructive feedback (which could be seen by everyone) on each observation with a misidentified insect or that appeared to violate the protocol. These comments aimed to give participants some tips to better identify insects and to better follow the protocol. As a result, participants soon started to critically assess newly uploaded contributions, leaving comments to remind authors of "suspicious" contributions about the standardized protocol and to explain the importance of abiding with it. This eventually led to a self-managed community that likely contributed to participants quickly learning the importance of following the standardized protocol for the sake of scientific research.

For the K-Spipoll, the only driver of respect of the protocol was the explanations of the researcher during the on-field training activities. Previous studies showed that an appropriate training of the participants with a professional scientist could be seen as one of the most important factors affecting their accuracy (Newman, Buesching, and Macdonald 2003; Silvertown et al. 2013). These events are important and allow exchanges between the observers and the researchers who are the recipients of the data. This direct contact can create a strong link between the scientists and the observers who can, in this way, better understand the stakes of their participation for biodiversity conservation and why respecting the protocol is scientifically important. The understanding of the scientific background has been shown to enhance participant's motivation and comprehension (Martinich, Solarz, and Lyons 2006). It is also a way for the researcher to share his/her knowledge and passion about a specific species or group of species and to make the observers want to participate. We suspect that the few training sessions organized the first year for the K-Spipoll were not attended by enough participants; additionally, attending a single training event might not be sufficient to ensure a full understanding of standardized research protocols.

These educational activities have been a way for a lot of participants to receive? experiences of nature in urban areas and to raise awareness about the importance of

pollinators for the functioning of ecosystems. Such "routine experiences of nature in cities" has been shown to increase personal commitment toward biodiversity conservation (Prévot et al. 2018).

#### *The website vs. the phone digital application*

The development of a cellphone digital application for the K-Spipoll presented advantages, decreasing the cost of participation by facilitating data entry. It might be the principal driver of participant engagement. The high number of sessions per observers for the K-Spipoll (13.3 against 2.8 for the Spipoll) and the greater spatial distribution of the sessions suggest that having the opportunity to participate at any time and anywhere with a smartphone and sending directly the observations could motivate the observers to participate more (although it can lead to a decrease of data quality, as mentioned above).

In the first two to three months of the Spipoll's start, its website encountered several bugs and crashes. This could have discouraged observers to upload their data, explaining the high number of participants who participated only once and the even greater number who "registered" on the website but never actually sent data. However, there was considerable effort from the Spipoll team to answer participant's questions about how to proceed with data uploading. Thus, website issues in initial months would not solely explain these participation patterns, which are more likely the result of the time necessary to participate.

#### *Pre-sorting of data and insect identification*

From the researchers' point of view, the organization of the Spipoll is more efficient, as only a validation by experts is required prior to data analyses: Participants carried out the time-consuming tasks of selecting the best picture of each insect recorded by session, and provided a first identification that was often correct, thanks to the online identification tool. For these observers, insect identification was a motivation to participate, bringing opportunities to learn more about pollinators and to improve their entomological skills (Deguines et al. 2018). Providing appropriate materials (e.g., online identification tools) to assist observers in insect identification, although challenging for pollinators, appears essential. However, learning to identify insects constitutes a demanding task that may discourage participants from continuing, explaining the high rate of one-time participation (60.8%).

In the case of the K-Spipoll, researchers had to find the best picture of each insect among the many photographs sent (including blurry or too-distant attempts). An additional substantial loss of time occurred as photographs lacked identification.

#### *Recommendations for the design of future programs*

Thanks to the transposition of a French citizen science program to South Korea, we were able to compare two very similar programs that nevertheless differed in a few characteristics. This unique opportunity allowed us to better understand the drivers influencing the quality of the data collected and participant engagement. Submission of observations via digital smartphone applications are becoming more popular in the field of citizen science (Liu et al. 2011; Newman et al. 2012; Land-Zandstra et al. 2016). The use of digital applications can also allow gamification (Tinati et al. 2017), which has the potential, thanks to a recreational and competitive approach, to recruit new participants by arousing their curiosity (Bowser et al. 2013) and to sustain engagement over time (Iacovides et al. 2013). The recruitment of sufficient participants every year and their commitment is critical to ensure the accuracy of data collection (especially for regular participants who are used to the protocol and who enhanced their identification skills), the collection of a large sample size every year, and the assessment of temporal dynamics of populations.

When Spipoll was launched in 2010, only 17% of the French population was equipped with smartphones (CREDOC 2016). Since then, their use increased dramatically: In 2016, 65% of the population possessed a smartphone (CREDOC 2016). In South Korea, 88% of the adults had a smartphone in 2016, which put the country at the highest smartphone ownership rate in the world (Pew Research Center 2016). By 2023, 3.5 billion persons may possess a smartphone (Ericsson Mobility report 2018). The development of digital applications on smartphones could thus be considered as a way to develop future citizen science programs. Smartphones can easily be used to collect data, thanks to all the tools integrated such as digital camera or microphones, which have been used to monitor treefrog habitat preferences in South Korea (Roh, Borzée, and Jang 2014). External devices can be used to improve the quality of the recording, such as ultrasonic microphones, used by the program iBat (Gibb, Mac, and Jones 2016).

However, to control the sending of accurate data, some features could be directly implemented on the application, such as protocol reminder questions ("Have you completed the required time of observation?" K-Spipoll has now been updated to ask this question in an attempt to improve the quality of the data); tick boxes to choose the best picture(s); and automatic identification allowing a first classification (i.e.,order and family).

## Conclusion

If the development of new technologies and digital applications can be seen as a convenient way to collect a large amount of data, implementation of controls at the stage of data collection is critical to ensure data quality and, therefore, the possibility to use these data to address ecological research questions. This paper showed that the process and methodology of the Spipoll program ensured that data collection was optimal for their analysis. This has been proven by the research papers published thanks to these data (Deguines et al. 2012; Deguines et al. 2016; Desaegher et al. 2018; Levé, Baudry, and Bessa-Gomes 2019).

The K-Spipoll process and methodology were more efficient to engage people to participate. The strong commitment of the observers is promising for the future of this program, for which data collection has been enhanced by adding controls into the application. Organization of on-field training sessions has been successful in engaging participants and providing experiences of nature in a

highly urban area, while meeting passionate researchers who can provide meaning to data collection.

Initial facilitation of a participant's network is also a key for later emergence of a self-organized community, where participants correct each other and share their skills and knowledge. It has been shown that the motivations of the observers can be linked to the sense of belonging to a social network while exchanging with people sharing the same interests (West and Pateman 2016; Domroese and Johnson 2017).

With this study, we highlighted how different methodologies between two similar pollinator monitoring programs led to various levels of data quality and participant's engagement, and we encourage researchers developing biodiversity monitoring programs relying on citizen science to carefully consider the multiple aspects presented here.

## Supplementary File

The supplementary file for this article can be found as follows:

- **Supplementary Materials.** Rationale for considering single-species sessions as suspicious. DOI: https://doi.org/10.5334/cstp.200.s1

## Acknowledgements

We would like to thank our partner Donga Science and all the participants from France and South Korea who participated in the two programs. We also want to warmly thank Dr. Lee Heung-Sik from Plant Quarantine Technology Center and Wan-Hyeok Park and Soo-Jeong Cho for their contributions to pollinator identification and the several hours they passed in the lab. We thank Ariel Jacobs for accurate review of the language. We are grateful to the editor and the three anonymous reviewers whose constructive comments improved our manuscript.

## Funding Information

This work was supported by the Korea Research Fellowship Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (KRF project grant number: 2016H1D3A1938095). ND was funded by the BiodivERsA3-2015-104 (BIOVEINS) grant.

## Competing Interests

The authors have no competing interests to declare.

## Author Contributions

All authors made substantial contributions to the conception and design of the work. HS, ND, and GL made substantial contributions to the acquisition of data. All authors contributed to the analysis and interpretation of the data. HS led the writing of the manuscript. ND, YJ, GL, and RJ critically contributed to the final version of the manuscript. All authors gave final approval for publication.

## References

**Bird, TJ, Bates, AE, Lefcheck, JS, Hill, NA, Thomson, RJ, Edgar, GJ, Stuart-Smith, RD, Wotherspoon, S, Krkosek, M, Stuart-Smith, JF, Pecl, GT, Barrett, N** and **Frusher, S.** 2014. Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173: 144–154. DOI: https://doi.org/10.1016/j.biocon.2013.07.037

**Bonney, R, Ballard, H, Jordan, R, McCallie, E, Phillips, T, Shirk, J** and **Wilderman, CC.** 2009. Public participation in scientific research: Defining the field and assessing its potential for informal science education. In: *Center for Advancement of Informal Science Education (CAISE) A CAISE Inquiry Group Report*. Washington, D.C.

**Bowser, A, Hansen, D, He, Y, Boston, C, Reid, M, Gunnell, L** and **Preece, J.** 2013. Using gamification to inspire new citizen science volunteers. In: *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, Toronto, Ontario, Canada, ACM, pp. 18–25. DOI: https://doi.org/10.1145/2583008.2583011

**Bruyere, B** and **Rappe, S.** 2007. Identifying the motivations of environmental volunteers. *Journal of Environmental Planning and Management*, 50(4): 503–516. DOI: https://doi.org/10.1080/09640560701402034

**Chandler, M, See, L, Buesching, CD, Cousins, JA, Gillies, C, Kays, RW, Newman, C, Pereira, HM** and **Tiago, P.** 2017. Involving Citizen Scientists in Biodiversity Observation. In: Walters, M and Scholes, R (eds.), *The GEO Handbook on Biodiversity Observation Networks*. The Netherlands: Springer. pp. 211–237. DOI: https://doi.org/10.1007/978-3-319-27288-7_9

**Chandler, M, See, L, Copas, K, Bonde, AMZ, López, BC, Danielsen, F, Legind, JK, Masinde, S, Miller-Rushing, AJ, Newman, G, Rosemartin, A** and **Turak, E.** 2016. Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213(B): 280–294. DOI: https://doi.org/10.1016/j.biocon.2016.09.004

**Cohn, JP.** 2008. Citizen Science: Can Volunteers Do Real Research? *BiosScience*, 53(3): 192–197. DOI: https://doi.org/10.1641/B580303

**Couvet, D, Jiguet, F, Julliard, R, Levrel, H** and **Teyssedre, A.** 2008. Enhancing citizen contributions to biodiversity science and public policy. *Interdisciplinary Science Reviews*, 33: 95–103. DOI: https://doi.org/10.1179/030801808X260031

**CREDOC.** 2016. Le Baromètre du Numérique (Edition 2016), n°R333. https://www.arcep.fr/uploads/tx_gspublication/presentation-barometre-du-numerique-291116.pdf.

**Curtis, V.** 2018. Patterns of Participation and Motivation in Folding@home: The Contribution of Hardware Enthusiasts and Overclockers. *Citizen Science: Theory and Practice*, 3(1): 5. DOI: https://doi.org/10.5334/cstp.109

**Deguines, N, de Flores, M, Loïs, G, Julliard, R** and **Fontaine, C.** 2018. Fostering close encounters of the entomological kind. *Frontiers in Ecology and the Environment*, 16(4): 202–203. DOI: https://doi.org/10.1002/fee.1795

**Deguines, N, Julliard, R, de Flores, M** and **Fontaine, C.** 2012. The whereabouts of flower visitors: contrasting land-use preferences revealed by a coun-

try-wide survey based on citizen science. *PLoS One*, 7(9): e45822. DOI: https://doi.org/10.1371/journal.pone.0045822

**Deguines, N, Julliard, R, de Flores, M** and **Fontaine, C.** 2016. Functional homogenization of flower visitor communities with urbanization. *Ecology and Evolution*, 6(7): 1967–1976. DOI: https://doi.org/10.1002/ece3.2009

**Desaegher, J, Nadot, S, Fontaine, C** and **Colas, B.** 2018. Floral morphology as the main driver of flower-feeding insect occurrences in the Paris region. *Urban ecosystems*. 21(4): 585–598. DOI: https://doi.org/10.1007/s11252-018-0759-5

**Devictor, V, Whittaker, RJ** and **Beltrame, C.** 2010. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions*, 16(3): 354–362. DOI: https://doi.org/10.1111/j.1472-4642.2009.00615.x

**Dickinson, JL, Zuckerberg, B** and **Bonter, DN.** 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefit. *Annual Review of Ecology, Evolution, and Systematics*, 41: 149–72. DOI: https://doi.org/10.1146/annurev-ecolsys-102209-144636

**Domroese, MC** and **Johnson, EA.** 2017. Why watch bees? Motivations of citizen science volunteers in the Great Pollinator Project. *Biological Conservation*, 208: 40–47. DOI: https://doi.org/10.1016/j.biocon.2016.08.020

**Ericsson Mobility Report.** 2018. https://www.ericsson.com/assets/local/mobility-report/documents/2018/ericsson-mobility-report-june-2018.pdf.

**Freitag, A, Meyer, R** and **Whiteman, L.** 2016. Strategies Employed by Citizen Science Programs to Increase the Credibility of Their Data. *Citizen Science: Theory and Practice*, 1(1): 1–11. DOI: https://doi.org/10.5334/cstp.6

**French Biodiversity Observatory.** 2018. http://indicateurs-biodiversite.naturefrance.fr/fr/indicateurs/evolution-de-limplication-des-citoyens-dans-les-sciences-participatives-liees-a-la.

**Gibb, R, Mac, O** and **Jones, K.** 2016. Bat Detective: citizen science for eco-acoustic biodiversity monitoring. *Environmental Scientist*, 25(2): 15–18.

**Guiney, MS** and **Oberhauser, KS.** 2009. Conservation Volunteers' Connection to Nature. *Ecopsychology*, 1(4): 187–197. DOI: https://doi.org/10.1089/eco.2009.0030

**Iacovides, I, Jennett, C, Cornish-Trestrail, C** and **Cox, AL.** 2013. Do games attract or sustain engagement in citizen science?: a study of volunteer motivations. *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, pp. 1101–1106. DOI: https://doi.org/10.1145/2468356.2468553

**Jiguet, F, Devictor, V, Julliard, R** and **Couvet, D.** 2012. French citizens monitoring ordinary birds provide tools for conservation and ecological sciences. *Acta Oecologica*, 44: 58–66. DOI: https://doi.org/10.1016/j.actao.2011.05.003

**Land-Zandstra, AM, Devilee, JL, Snik, F, Buurmeijer, F** and **van den Broek, JM.** 2016. Citizen science on a smartphone: Participants' motivations and learning. *Public Understanding of Science*, 25(1): 45–60. DOI: https://doi.org/10.1177/0963662515602406

**LeBuhn, G, Connor, E, Brand, M, Coville, J, Devkota, K, Thapa, R, Kasina, M, Joshi, R, Aidoo, K, Kwapong, P, Annoh, C, Bosu, P** and **Rafique, M.** 2016. Monitoring Pollinators Around the World. In: Gemmill-Herren, B (ed.), *Pollination Services to Agriculture*. In: Gemmill-Herren, B (ed.). London: Routledge.

**Levé, M, Baudry, E** and **Bessa-Gomes, C.** 2019. Domestic gardens as favorable pollinator habitats in impervious landscapes. *Science of the Total Environment*, 647(10): 420–430. DOI: https://doi.org/10.1016/j.scitotenv.2018.07.310

**Lewandowski, E** and **Specht, H.** 2015. Influence of volunteer and project characteristics on data quality of biological surveys. *Biological Conservation*, 29(3): 713–723. DOI: https://doi.org/10.1111/cobi.12481

**Lewandowski, EJ** and **Oberhauser, KS.** 2017. Butterfly citizen scientists in the United States increase their engagement in conservation. *Biological Conservation*, 208: 106–112. DOI: https://doi.org/10.1016/j.biocon.2015.07.029

**Liu, Y, Piyawongwisal, P, Handa, S, Yu, L, Xu, Y** and **Samuel, A.** 2011. Going Beyond Citizen Data Collection with Mapster: A Mobile + Cloud Real-Time Citizen Science Experiment. 2011. In: *IEEE Seventh International Conference on e-Science Workshops*, Stockholm, pp. 1–6. DOI: https://doi.org/10.1109/eScienceW.2011.23

**Martinich, JA, Solarz, SL** and **Lyons, JR.** 2006. Preparing students for conservation careers through projectbased learning. *Conservation Biology*, 20: 1579–1583. DOI: https://doi.org/10.1111/j.1523-1739.2006.00569.x

**Muratet, A** and **Fontaine, B.** 2015. Contrasting impacts of pesticides on butterflies and bumblebees in private gardens in France. *Biological Conservation*, 182: 148–154. DOI: https://doi.org/10.1016/j.biocon.2014.11.045

**Newman, C, Buesching, CD** and **Macdonald, DW.** 2003. Validating mammal monitoring methods and assessing the performance of volunteers in wildlife conservation – "Sed quis custodiet ipsos custodies?" *Biological Conservation*, 113: 189–197. DOI: https://doi.org/10.1016/S0006-3207(02)00374-9

**Newman, G, Wiggins, A, Crall, A, Graham, E, Newman, S** and **Crowston, K.** 2012. The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, 10(6): 298–304. DOI: https://doi.org/10.1890/110294

**Olivier, T, Schumki, R, Fontaine, B, Villemey, A** and **Archaux, F.** 2016. Butterfly assemblages in residential gardens are driven by species habitat preference and mobility. *Landscape Ecology*, 31(4): 865–876. DOI: https://doi.org/10.1007/s10980-015-0299-9

**Ollerton, J, Winfree, R** and **Tarrant, S.** 2011. How many flowering plants are pollinated by animals? *Oikos*, 120(3): 321–326. DOI: https://doi.org/10.1111/j.1600-0706.2010.18644.x

**Pew Research Center.** 2016. Smartphone Ownership and Internet Usage Continues to Climb in

Emerging Economies. http://assets.pewresearch.org/wp-content/uploads/sites/2/2016/02/pew_research_center_global_technology_report_final_february_22__2016.pdf.

**Ponciano, L** and **Brasileiro, F.** 2014, Finding Volunteers' Engagement Profiles in Human Computation for Citizen Science Projects. *Human Computation*, 1(2): 247–266. DOI: https://doi.org/10.15346/hc.v1i2.12

**Potts, SG, Biesmeijer, JC, Kremen, C, Neumann, P, Schweiger, O** and **Kunin, WE.** 2010. Global pollinator declines: trends, impacts and drivers. *Trends in Ecology and Evolution*, 25(6): 345–353. DOI: https://doi.org/10.1016/j.tree.2010.01.007

**Prévot, AC, Cheval, H, Raymond, R** and **Cosquer, A.** 2018. Routine experiences of nature in cities can increase personal commitment toward biodiversity conservation. *Biological Conservation*, 226: 1–8. DOI: https://doi.org/10.1016/j.biocon.2018.07.008

**Ries, L** and **Oberhauser, K.** 2015. A Citizen Army for Science: Quantifying the Contributions of Citizen Scientists to our Understanding of Monarch Butterfly Biology. *BioScience*, 65(4): 419–430. DOI: https://doi.org/10.1093/biosci/biv011

**Roh, G, Borzée, A** and **Jang, Y.** 2014. Spatiotemporal distributions and habitat characteristics of the endangered treefrog, *Hyla suweonensis*, in relation to sympatric *H. japonica*. *Ecological Informatics*, 24: 78–84. DOI: https://doi.org/10.1016/j.ecoinf.2014.07.009

**Silvertown, J, Buesching, CD, Jacobson, SK** and **Rebelo, T.** 2013. Citizen science and nature conservation. In: *Key Topics in Conservation Biology 2*, First Edition. In: Macdonald, DW and Willis, KJ (eds.). USA: John Wiley & Sons, pp. 127–142. DOI: https://doi.org/10.1002/9781118520178.ch8

**Suzuki-Ohno, Y, Yokoyama, J, Nakashizuka, T** and **Kawata, M.** 2017. Utilization of photographs taken by citizens for estimating bumblebee distributions. *Scientific Reports*, 7: 11215. DOI: https://doi.org/10.1038/s41598-017-10581-x

**Theobald, EJ, Ettinger, AK, Burgess, HK, DeBey, LB, Schmidt, NR, Froehlich, HE, Wagner, C, HilleRisLambers, J, Tewksbury, J, Harsch, MA** and **Parrish, JK.** 2015. Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation*, 181: 236–244. DOI: https://doi.org/10.1016/j.biocon.2014.10.021

**Thornhill, I, Loiselle, S, Lind, K** and **Ophof, D.** 2016. The Citizen Science Opportunity for Researchers and Agencies. *BioScience*, 66(9): 720–721. DOI: https://doi.org/10.1093/biosci/biw089

**Tinati, R, Luczak-Roesch, M, Simperl, E** and **Hall, W.** 2017. An investigation of player motivations in Eyewire, a gamified citizen science project. *Computers in Human Behavior*, 73: 527–540. DOI: https://doi.org/10.1016/j.chb.2016.12.074

**Trumbull, DJ, Bonney, R, Bascom, D** and **Cabral, A.** 2000. Thinking scientifically during participation in a citizen science project. *Science Education*, 84: 265–275. DOI: https://doi.org/10.1002/(SICI)1098-237X(200003)84:2<265::AID-SCE7>3.0.CO;2-5

**West, S** and **Pateman, R.** 2016. Recruiting and Retaining Participants in Citizen Science: What Can Be Learned from the Volunteering Literature? *Citizen Science: Theory and Practice*, 1(2): 1–10. DOI: https://doi.org/10.5334/cstp.8

**Wiggins, A, Newman, G, Stevenson, RD** and **Crowston, K.** 2011. Mechanisms for Data Quality and Validation in Citizen Science. In: *Seventh IEEE International Conference on e-Science Workshops*, pp. 14–19. DOI: https://doi.org/10.1109/eScienceW.2011.27

**How to cite this article:** Serret, H, Deguines, N, Jang, Y, Lois, G and Julliard, R. 2019. Data Quality and Participant Engagement in Citizen Science: Comparing Two Approaches for Monitoring Pollinators in France and South Korea. *Citizen Science: Theory and Practice*, 4(1): 22, pp. 1–12. DOI: https://doi.org/10.5334/cstp.200

**Submitted:** 07 September 2018 **Accepted:** 26 May 2019 **Published:** 18 July 2019

**Copyright:** © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See https://creativecommons.org/licenses/by/4.0/.

]u[ *Citizen Science: Theory and Practice* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS